MICROBIAL MARKER GENE INDICATORS FOR SOIL GREENHOUSE GAS EMISSIONS IN SOUTHEAST ALASKA

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Masters of Science

by

Sarah Ann Nadeau

December 2018

ABSTRACT

Soils in Southeast Alaska contain some of the world's highest concentrations of organic matter; however, the balance between carbon-sequestering net primary production and greenhouse gas (GHG)-emitting soil respiration in this region may shift due to climate change. We conducted a metagenomic survey of microbial respiratory marker gene abundances in soils from several ecosystem types in the carbon (C)-rich perhumid coastal temperate rainforest (PCTR) of Southeast Alaska in order to better understand microbial controls on GHG emissions in the context of a vulnerable soil C stock. Landscape-scale variability in aerobic respiration, methanogenesis, and denitrification marker genes across ecosystem types was most significantly related to gradients in soil moisture and pH. We found that the aerobic respiration marker gene A1-type heme-copper oxidase was associated with measured summertime carbon dioxide (CO_2) emissions and the methanogenesis marker gene *mcrA* was associated with summertime methane (CH_4) emissions. Denitrification marker genes *nirK*, *nirS*, *cNor*, *qNor*, and *nosZ* were not associated with measured nitrous oxide (N_2O) emissions; however, denitrification marker gene abundance was associated with denitrification activity as measured by denitrification enzyme assays. Denitrification activity was shown to be strongly controlled by nitrate rather than denitrifying community composition. This study demonstrates the utility of microbial respiratory marker genes in understanding where in the landscape the greatest emissions of the greenhouse gases CO_2, CH_4, and N_2O likely occur and what soil properties regulate microbially-mediated GHG emissions.

BIOGRAPHICAL SKETCH

Sarah Nadeau was born and raised in Rochester, NY with her three brothers and a rotating cast of extra family members and foreign exchange students. She graduated from Cornell University in Ithaca, NY with bachelors of science degrees in Plant Science and Biological Engineering in 2016. Sarah completed several co-op positions in plant breeding, moving across the U.S. to live and work in California and Iowa. Through her plant breeding experiences, Sarah developed an interest in data-driven problem solving and computational analysis. She returned to Cornell for a masters of science in Biological Engineering, during which she learned to code and applied herself to using genetic data to inform our understanding of soil nutrient cycling. In her free time, Sarah enjoys marathon running, cycling, travelling, and reading.

ACKNOWLEDGEMENTS

I would like to thank my committee members Todd Walter and Jim Shapleigh for their valuable support and advice. Thank you to Liz Kreitinger for being instrumental to the success of this study, Dave D'Amore for his comments on initial data, Christine Georgakakos for her help with sampling, and Leah Balkin, Clare O'Connor, and Sharon Zhang for their help with DNA extractions. I owe my peers in the Soil and Water Lab a debt of gratitude for their friendship. To my partner Sammy, thank you for being both a constant source of encouragement and a partner in adventure. Finally, thank you to my family - Peter, Molly, Patrick, Andy, Jane, Dave, and Nana - for making me who I am.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

To improve our understanding of the Southeast Alaskan perhumid coastal temperate rainforest (PCTR)'s carbon-rich soils as a current sink and potential future source of greenhouse gases (GHGs), we need to understand the role of soil microbial communities in this environment as GHG consumers and producers. Globally, soil microorganisms are critical influencers of the climate. In total, soils emit over 350 Pg of carbon dioxide (CO_2) equivalents per year of GHGs, mainly CO_2, CH_4, and N_2O [40]. Approximately 50% of total soil CO_2 emissions are believed to be produced by microbial respiration, and microbial respiration is the only major source of CH_4, and N_2O emissions from soils [40]. The PCTR is a particularly important setting in which to explore microbial controls on GHG gas cycling because this region's soils are some of the densest stocks of soil organic matter on the planet (>300 Mg Carbon (C) ha^{-1}, [15]). Release of this organic matter-bound C due to increased organic matter decomposition and coupled microbial respiratory activity could contribute large amounts of GHGs to the atmosphere [9]. Fortunately, PCTR soils are currently a net C sink, sequestering an estimated 1.0 to 1.8 Mg C ha^{-1} $year^{-1}$ as net primary production outpaces soil respiration [14]. However, climate change-driven alterations in temperature and precipitation patterns may significantly change the C sequestration versus C sink properties of PCTR soils by deferentially stimulating net primary production and soil respiration. The objective of this study is to investigate microbial controls on GHG emissions from PCTR soils by exploring

the utility of respiratory pathway marker gene abundance as indicators of GHG emissions.

Prior research in high-latitude tundra and boreal forests has provided evidence of mixed responses in both net primary production and soil respiration to climate change, depending on the region studied, the estimation approach employed, and modeling assumptions made [5, 9]. While tundra regions were generally reported to exhibit a greater increase in soil respiration than net primary production (e.g. [48]), in boreal forests increases in net primary production are thought to balance increases in soil respiration (e.g. [27]) [5]. While prior studies have emphasized the role of tundra and boreal forest zones as C sinks or sources, the PCTR represents a range of ecosystem types whose contributions to GHG emissions have been less well studied, yet are subject to similar climate perturbations. As in tundra and boreal forest regions, temperature and precipitation have both increased in the PCTR. 50-year temperature measurements (1949 - 1997) from Juneau and Yakutat, AK indicate that surface temperatures in the PCTR have warmed between 0.12 and 0.21 ^{o}C per 10 years [27]. Additionally, [54] report a 10% increase in annual precipitation for the South/Southeast region of Alaska over the same time period. Recently, [16] performed laboratory incubations of PCTR soils and found that increased temperature stimulated decomposition of soil organic matter and was positively correlated with increases in CO_2 flux. These results suggest that microbial community respiration exerts strong control over CO_2 fluxes in the PCTR and may be stimulated by increased temperature. Other studies in the Alaskan fen and UK peat bogs have also shown that microbial breakdown of soil organic matter and respiration is controlled by temperature (e.g. [6, 7]). These studies focus on soil CO_2 emissions as a driver of a global climate change. However, methane

(CH_4) and nitrous oxide (N_2O) are two other microbially-produced GHGs with significant implications for climate change, with 3.7 and 180 times the global warming potential of CO_2, respectively [33]. In this study, we extend the work of [16] in quantifying GHG emissions across a range of PCTR soils by considering fluxes of all three GHGs *in-situ*.

Beyond simply quantifying current GHG emissions from soils, determining future GHG contributions to the atmosphere from PCTR soils requires an understanding of the microbial controls on GHG-producing respiratory pathways. The extent to which a soil microbial community produces GHGs in soils depends on which GHG-producing respiratory pathways community members are genetically capable of employing and which are favored by local soil conditions. [45] investigated microbial marker genes for nitrogen (N)-cycling pathways in interior Alaska and found that marker gene abundance was associated with potential denitrification and nitrification rates. These promising results indicate that marker gene abundance may be a useful indicator of soil respiratory activity. In this study, we employed a similar marker gene-survey strategy in the PCTR, but extend the work of [45] to include marker genes from CO_2 and CH_4-producing respiratory pathways. We additionally link marker gene abundance to *in-situ* GHG fluxes as well as potential denitrification activity. To more broadly investigate microbial controls on GHG emissions in the PCTR, this study seeks to address the following two main questions: (1) Do microbial communities across PCTR ecosystem types vary in their genetic capacity to perform aerobic respiration, methanogenesis, and denitrification and (2) If so, are variations in functional gene abundance associated with variations in GHG emissions? Finally, because we made measurements of both *in-situ* N_2O emissions and potential denitrification activity under different conditions, we

were able to ask a third question pertaining to controls on denitrification activity specifically: to what extent does total denitrification gene abundance moderate denitrification activity in PCTR soils?

In order to evaluate microbial controls on CO_2, CH_4, and N_2O emissions across a range of PCTR soils, we quantified functional marker gene abundance from three different respiratory pathways - aerobic respiration, methanogenesis, and denitrification - across several PCTR ecosystem types using metagenomic sampling. These respiratory pathways were chosen because each produces a GHG; aerobic respiration produces CO_2, methanogenesis produces CH_4, and denitrification produces N_2O as an intermediate product. Six PCTR ecosystem types were chosen for sampling to incorporate a range of PCTR soil conditions. The upland, palustrine forested wetland, palustrine scrub-shrub wetland, and palustrine emergent wetland ecosystem types were chosen because they span a hydrologic gradient (from driest to wettest) along hillslopes and are common across the PCTR. Finally, riparian areas without salmon and with salmon were additionally considered in order to investigate the impact of annual influxes of nitrogen (N) and C-rich organic matter from spawning salmon in some PCTR streams.

1.2 Advantages and Limitations of the Metagenomics Approach

The metagenomic DNA sequencing approach to quantifying functional gene abundance has unique advantages and limitations as a means of investigating microbial community structure-soil process links. Briefly, metagenomic DNA

sequencing involves the lysis of cells from a small soil volume, the genomes of which are broken into small pieces and sequenced. Resulting DNA sequences, also known as reads, are short (150 base pairs in this study) and come from a multitude of different genomes. Depending on sequencing depth, many genomes from the original soil sample may not be represented fully, if at all, in the resulting collection of metagenomic reads. As a result, it is often difficult to re-assemble longer portions of original genomes from collections of metagenomic reads. Since metagenomic reads are short, difficult to assemble into longer genome fragments, and originate from potentially highly diverse organisms, assigning taxonomy to reads is difficult. Furthermore, the genome databases which taxonomy assignment methods rely on represent only a small fraction of earths total diversity, much of which is not taxonomically annotated [35]. Limited sequencing depth and difficulty in taxonomically classifying reads are the main drawbacks to metagenomic sampling.

Despite these limitations, metagenomic sampling has several advantages over the alternate amplification-based (qPCR) approach for quantifying functional gene abundance. These advantages made a metagenomic survey the most suitable approach to answer our study questions. Many previous studies have used qPCR as a means of quantifying nutrient cycling gene abundance in soils (e.g. [45] in Alaska, see [34] for a review). During qPCR functional gene quantification, barcode primer sequences are used to select and amplify only genes of interest from a soil sample. The limited scope of qPCR sequencing results makes phylogenetic analysis and taxonomic assignment relatively easier than from metagenomic sequencing. However, qPCR is subject to primer bias; sequence amplification depends on highly specific matches to the selected primer(s). Whereas metagenomic sampling might result in a wide diversity of

identified gene sequences, qPCR results are limited to sequences closely related to primer sequences. Primer bias is known to result in underestimates of gene abundance from phylogenetically diverse sequences [56]. For example, primer bias was shown to result in significant underestimates of some *nirK* sequences in one denitrification study [10]. The choice of qPCR or metagenomic DNA sequencing includes a trade-off between accurate gene quantification and taxonomic resolution of sampling. Since this study aims to quantify the relative abundance of a multitude of marker genes from potentially diverse soil microbial communities, accurate gene assignment and unbiased gene quantification were priorities. As a result, we chose to perform metagenomic sampling at the expense of taxonomic resolution for this study.

1.3 Respiratory Pathways & Marker Genes

In this study, marker genes from three GHG-producing microbial respiratory pathways are considered - aerobic respiration, methanogenesis, and denitrification. Briefly, heme-copper oxidase (HCO) types are used as marker genes for aerobic respiration, *mcrA* is used as a marker gene for methanogenesis, and the genes *nap*, *nar*, *nirK*, *nirS*, *cNor*, *qNor*, and *nosZ* are used as marker genes for denitrification.

Aerobic respiration is the most energetically favorable respiratory pathway for soil microorganisms and is carried out by the majority of bacteria and archaea found in soils. During aerobic respiration, carbon sources are oxidized, producing CO_2, and the reducing equivalents are used by a membrane associated respiratory chain to make ATP. Crucially, oxygen (O_2) is required as the

terminal electron acceptor for aerobic respiration. Aerobic respiration in soils generally increases with increasing microbial biomass, soil water content (to a point), temperature, and soil nutrient content [46, 60]. Total soil CO_2 fluxes include respiratory contributions from plant roots, soil fauna, and soil microorganisms. Of these CO_2 sources in soils, the relative contribution of microorganisms varies greatly, though it is approximated to be about 50% of total soil respiration [24]. Regardless of the exact contribution, microbial respiration is certainly a significant contributor to atmospheric CO_2 levels [40, 20].

O_2 reductases catalyze the reduction of O_2 to water as the final step in aerobic respiration. The most prevalent O_2 reductases are those belonging to the superfamily of heme-copper oxidases (HCOs) [43]. HCOs are typically divided into at least 4 classes based on phylogenetic and structural differences, with additional subgroups for some classes. The nomenclature introduced by [43] and supported in a subsequently expanded analysis by [51] defines these classes as A-, B-, and C-type HCOs, with A-type oxidases further subdivided into A1- and A2-types based on a differing amino acid motif at one location. The structural differences among the various HCO types give rise to differing oxygen affinities and energetic benefits. A-type HCOs are known as low-O_2 affinity with high energetic benefit, while C-type HCOs are known as high-O_2 affinity with low energetic benefit [23].

In anaerobic conditions, some organisms are able to reduce CO_2 during respiration, a process called methanogenesis. Methanogenesis is carried out by strictly anaerobic archaea. During methanogenesis, acetate, or H_2 and CO_2, are reduced to methane gas (CH_4). Methanogenesis is favored in extremely low-O_2 environments where NO_3^- and other more energetically favorable alternate

terminal electron acceptors like sulfate (SO_4^{2-}), manganese (Mn(IV)), and iron (Fe(III)) are limited [57]. The methyl-coenzyme M reductase (*mcrA*) enzyme catalyzes the final step in methanogenesis, converting methyl-CoM into CH_4. This enzyme is common to all known methanogens and is typically used as a marker gene for methanogenesis [12]. Methanogenesis is the sole source of soil CH_4 emissions [11].

Additionally, nitrate (NO_3^-) can be used as the terminal electron acceptor through two different respiratory pathways: denitrification and dissimilatory nitrate reduction to ammonium (DNRA). During denitrification, NO_3^- is sequentially reduced to nitrite (NO_2-), nitric oxide (NO), nitrous oxide (N_2O), and finally dinitrogen gas (N_2) by the *nap*, *nar*, *nirK*/*nirS*, *cNor*/*qNor*, and *nosZ* enzymes, respectively. However, many denitrifying microorganisms are capable of only one or a few of these reactions, so the pathway may proceed in an incomplete or step-wise fashion with multiple microorganisms involved. Denitrifying microorganisms are a phylogenetically-diverse group of facultatively-anaerobic bacteria and archaea. It is generally accepted that denitrification is favored in oxygen-limited (high soil water content) soils with abundant NO_3^- and labile C [18]. During DNRA, on the other hand, NO_3^- is reduced to ammonium (NH_4^+). While there is some evidence N_2O may be released as a byproduct of DNRA, this amount is generally considered negligible compared to N_2O emissions from denitrification [31]. Denitrification is considered the primary source of N_2O emissions from soils [34].

1.4 Study Sites

The PCTR of Southeast Alaska covers approximately 8.5 million ha along the western coast of North America, extending from Yakutat to Ketchikan, Alaska [15, 13]. Precipitation is frequent and year-round, averaging 1,500 to 5,000 mm annually [15]. Temperatures are generally cool, ranging from an average temperature of 28.3 °F in January to an average temperature of 55.9 °F in August at the Juneau Airport weather station (1981-2010 NOAA Climate Normals). Cool temperatures and moist conditions limit nutrient turnover, giving rise to low SOC turnover and large soil C stocks (>300 Mg C ha^{-1}, [15]). Atmospheric N deposition is low (4.96 kg N $ha^{-1}yr^{-1}$, [25]) and PCTR soils are characterized as N-limited [17]. Hydrologic conditions strongly control soil conditions in the PCTR, giving rise to differentiated plant communities along hydrologic gradients [3]. In this study, the same ecosystem types used by Bisbing & D'Amore ([3, 2]) are investigated because they span a range of hydrologic conditions and are ubiquitous across the PCTR region. Wetlands, represented by four of the ecosystem types considered in this study, comprise approximately 20% of PCTR land area [16]. The ecosystem types used by Bisbing & D'Amore ([3, 2]) are categorized based on predominant flora according to the National Wetland Inventory (NWI) classification system [8]. These ecosystem types include the palustrine emergent wetland (PEM, emergent wetland), the palustrine scrubshrub wetland (PSS, scrub-shrub wetland), the palustrine forested wetland (PFO, forested wetland), and the forested upland (U, upland). The palustrine emergent wetland is characterized by emergent herbaceous angiosperms, while the palustrine scrub-shrub wetland is characterized by shrubs and small trees [3]. According to the extensive site hydrology characterization performed

by Bisbing & D'Amore in [2], both ecosystem types experience near-constant soil saturation, with depths to water table ranging from 3-8 cm for emergent wetland soils and 4-15 cm for scrub-shrub wetland soils. The forested wetland ecosystem type supports large trees, though soils remain saturated for approximately half the growing season. Mean annual depth to groundwater in forested wetland soils ranges from 15-63 cm. Finally, the forested upland ecosystem type of the PCTR is characterized by mixed-conifer forests and a mean depth to water table in the range of 79-96 cm. Depth to groundwater rarely, if ever, encroaches into the 20 cm rooting zone in upland soils.

In addition to the hydrologically-differentiated emergent wetland, scrub-shrub, forested wetland, and upland ecosystem types, this study also considers riparian sites both accessible to salmon (RF, riparian with salmon) and inaccessible to salmon (RNF, riparian without salmon). The annual run of spawning salmon brings rich N and C resources from marine environments to inland riparian areas where the fish reproduce and then senesce. These nutrient inputs are hypothesized to significantly stimulate the local ecosystem (see review by [26]). Outside stream banks, several studies have shown that salmon carcasses distributed to stream banks by bears could have significant impacts on localized soil N and C pools [19, 59]. In order to investigate microbial respiratory activity in the context of salmon-derived N and C inputs, riparian sites accessible and inaccessible to salmon were included in this study.

Three sites of the same soil type within each of the upland, forested wetland, scrub-shrub wetland, emergent wetland, and riparian with salmon ecosystem types and two sites of the same soil type within the riparian without salmon ecosystem type were chosen to encompass ecosystem type heterogeneity while

controlling for physical soil characteristics. Only two sites were available for the riparian without salmon ecosystem type due the scarcity of riparian sites inaccessible to fish (above waterfalls), accessible to people, and with the same soil type. Within each site, soil was sampled at 4 replicate plots. A total of 68 plots were sampled, with sampling design blocked by both site and ecosystem type. Ecosystem types are ordered in figures from driest to wettest (U, PFO, PSS, PEM) and then riparian without and with salmon (RNF, RF).

CHAPTER 2

METHODS

2.1 Soil Property Measurements

At each plot, two soil cubes, roughly 10 cm on a side, were sampled and composited for measurement of soil properties. From the composited soil, two replicate samples of 6 ± 0.5 g soil were used for KCl extractions. KCl extractions were performed at the Cary Institute of Ecosystem Studies (CIES) according to the protocol detailed in [47]. 30 mL of 1M KCl was added to soil samples, which were vortexed, shaken, and centrifuged. After centrifugation, soluble nitrate (NO_3^-), ammonium (NH_4^+), and phosphorus (P) were quantified in the supernatant using a Lachat ion chromatograph (detection limit of 0.02 mg/L). Additionally, two replicate samples of 15 ± 0.5 g soil were mixed with deionized water to form a soil/water slurry in which pH was measured. Finally, two replicates samples of 7.5 ± 0.5 g of soil were dried and weighed to calculate gravimetric water content (GWC).

2.2 GHG Chamber Flux Measurements

In-situ fluxes of CO_2, CH_4, and N_2O ("gases") were measured from two plots within each site in July 2017. Measurements were replicated during two sampling campaigns conducted several weeks apart to determine how results varied intra-seasonally. Measurements were made using the static chamber technique described in [42], in which a circular collar is installed below the soils

surface and a cover is added to create an airtight chamber above a small area of soil. Soil collars were installed one week prior to first sampling to allow soils adequate time to equilibrate. These chambers were then used to capture trace gases produced by the enclosed soil over a period of 30 minutes. Chamber soil coverage area and total enclosed volume were 0.07 m^2 and 0.02 m^3, respectively. During sampling, accumulated gases within the chamber were sampled using a syringe and transferred to evacuated glass vials at 0, 10, 20, and 30 minutes after the placement of chamber covers. Gas concentrations in the sample vials were determined by gas chromatography at CIES. Gas fluxes were calculated by fitting a linear model to the concentration data from each chamber over time. Data from chambers for which linear model fits for CO_2 concentration were very poor (R^2 value of less than 0.7, 17 chambers) or in which CO_2 flux was negative (2 additional chambers) were removed since these symptoms indicate that the chamber seal may have been broken during sampling. In total, 49 valid chamber measurements remained after these data quality control steps. While no negative gas concentrations were calculated from valid chamber measurements, several concentration values were outside the bounds of the gas standards used to calibrate gas chromatography measurements. Namely, 1 N_2O concentration was reported to be 0.8 times the lowest N_2O standard (0.25 ppm) and 12 CH_4 concentrations were reported to be higher than the highest CH_4 standard (10.1 ppm). One soil plot in particular produced exceedingly high CH_4-concentration samples, with peak CH_4 concentrations reaching 26 times that of the highest CH_4 standard. The remainder of the extraordinarily high CH_4-concentration samples were all within 5 times the concentration of the highest CH_4 standard. A paired t-test was performed using flux data from plots for which valid chamber gas flux measurements were achieved in both sampling campaigns (n = 27

plots), which revealed that measured CO_2 fluxes were on average 46.4 mg m^{-2} hr^{-1} higher during the second sampling campaign ($p = 0.005$). There was no statistically significant difference in measured CH_4 or N_2O fluxes between the two campaigns ($p > 0.05$). For soil plots with valid flux measurements from both campaigns, fluxes were averaged and otherwise the measurement from only the campaign with a valid measurement was reported. In ANOVA tests to determine significant differences in gas fluxes between ecosystem types a categorical 'campaign' variable was included as a random effect.

2.3 Denitrification Enzyme Assays

In order to quantify denitrification rates, denitrification enzyme assays (DEAs) were performed. Since N_2, rather than N_2O, is the final end product of denitrification, chamber flux measurements of N_2O did not measure denitrification rates, per se. Instead, denitrification rates were quantified with DEA experiments. During DEA experiments, anoxic conditions were induced in soil cores by flushing cores with N_2 gas. Acetylene gas (C_2H_2) was then added to inhibit N_2O conversion to N_2. As a result, the denitrification process is terminated with N_2O as the final end product. N_2O flux over a 60-minute period (from 30 to 90 minutes after the start of incubation) was then quantified via gas chromatography, also performed at CIES. The complete DEA methodology is described in [21]. Since DEA terminates denitrification with N_2O as a final product, N_2O fluxes from DEA experiments are assumed to be representative of relative denitrification activity. Three DEA experiments were performed: one incubation with soils only and no added nutrients (soil-only treatment), and two replicate experiments with soils plus added NO_3^- and glucose (soil + NO_3^- + glucose treat-

ment). For all DEA experiments, two replicates of 5 ± 0.5 g soil were used for the incubation. The DEA media for the soil + NO_3^- + glucose treatment was 10 mL of 100 mg/L NO_3^-N and 1000 mg/L glucose. N_2O concentrations and fluxes were determined in the same manner as for chamber gas flux measurements, though N_2O concentration was only sampled at two timepoints, 30 and 90 minutes. It is worth noting that these results must be interpreted with caution since measured N_2O concentrations were frequently lower than the lowest N_2O standard used to calibrate gas chromatography results (0.25 ppm). Across samples from all three DEA experiments, reported N_2O concentrations were lower than the lowest N_2O standard for 148 samples out of 408. Additionally, 17 samples were reported to have negative concentrations, a unphysical result which was accounted for by adjusting up to zero. No N_2O measurements were discarded from the study. N_2O fluxes varied by 4.8 μg N g soil^{-1} hr^{-1} on average between the two soil + NO_3^- + glucose experiments (paired t-test, p = 0.02). To account for these differences, N_2O fluxes for the soil + NO_3^- + glucose treatment were reported as an average between the two replicate experiments. In ANOVA tests to determine significant differences in DEA results between ecosystem types a categorical 'replicate' variable was included as a random effect for the soil + NO_3^- + glucose experiments.

2.4 Metagenomic Sampling, DNA Extraction, & Sequencing

Metagenomic sampling was performed at the plot level (the finest resolution sampling level in this experiment), resulting in a total of 68 samples with 8-12 samples per ecosystem type. First, three soil cubes, roughly 10 cm on a side, were taken from each plot. To limit contamination during soil collection, cubes

were broken open and soil was sampled from the center of each cube. Samples from the center of each of the three cubes were composited and immediately placed in sterile containers. Soils were then frozen and overnight shipped to Cornell University in Ithaca, NY. Between 24 and 125 g of soil were sampled in total from each plot. DNA was extracted from 0.25 g of soil per sample using a MoBio PowerSoil kit and quantified using a Qubit fluorometer. Library preparation and sequencing were performed by the Biotechnology Resource Center (BRC) at the Cornell Institute of Biotechnology. Library preparation was done with a NexteraXT kit and single-end 150 base pair DNA sequencing was performed on an Illumina NextSeq500 machine. Remaining Nextera sequencing adapters were removed using Cutadapt with default settings [37] and read quality for the resulting sequencing files was assessed using FastQC [1]. One metagenomic sample (PEM site 2, plot A) was removed from the analysis due to a very low number of reads (<2 standard deviations below the average number of reads across samples). For remaining, quality-controlled samples, the number of reads per sample ranged from 2,451,038 to 9,485,771 with an average of 4,157,109 reads per sample. A breakdown of raw and quality-controlled read counts per sample is available in Appendix B Table 2.

2.5 Functional Gene Assignment

Functional gene assignment was performed using a similarity-based approach in which reads are translated from DNA to protein sequences and best-match alignments are found with annotated reference protein databases. To investigate broad trends in overall community genetic structure, all reads were first aligned to a filtered Uniref90 dataset using FMAP [29]. The Uniref90 dataset includes

protein sequences from the Uniref databases clustered at 90% sequence identity. It has been shown that 98% of Uniref90 clusters contain proteins of identical function, making this dataset useful for functional assignment [55]. Additionally, the Uniref90 dataset includes information on the common taxonomy of each cluster. To reduce spurious gene assignments, we utilized this Uniref90 feature to filter the dataset to include only sequences classified as bacterial (NCBI taxonomy ID 2) and archaeal (NCBI taxonomy ID 2157). The resulting filtered database included 3,849,454 protein sequences from 13,481 KEGG orthologous groups ("orthologies"), where an orthology encompasses a collection of homologous gene sequences. We then ran performed similarity-based functional annotation of metagenomic reads with the default maximum e-value of 10. Between 4 and 8% of reads from each sample were assigned to one of 10,286 KEGG orthologies. The breakdown of percentage reads assigned a function is available in Appendix B Table 2. We chose to report functional annotation results as reads per kilobase per million (RPKM) for each orthology because RPKM values are crudely normalized to both gene length and sequencing depth (see [29] for a detailed description of RPKM calculation). By reporting orthology abundance as RPKM values, we aim to prevent the underestimation of very short genes and overestimation of very long genes in our samples.

To get a more accurate estimate of abundance for respiratory pathway marker genes, all reads were additionally aligned to manually curated protein sequence databases for each selected marker gene. 1,781 HCO protein sequences were downloaded from the database compiled by Sousa *et al.* ([50]), 7,514 denitrification protein sequences were downloaded from the database compiled by Li *et al.* ([36]), and 1,460 *mcrA* protein sequences were downloaded from the database compiled by Speth & Orphan ([53]). The break-

down of sequences from each protein or protein type within these databases is detailed in Appendix B Table 1. We chose to perform functional assignment for marker genes using individually curated databases, rather than the more general Uniref90 database discussed above because we wanted to encompass as much of the known diversity in marker gene sequences as possible. To avoid spurious gene assignments and alignment to distant homologs of marker genes, we used a stricter maximum e-value of 0.001 for these alignments. Reads were aligned to each database using DIAMOND [4]. Read-database sequence matches were then filtered to include only alignments with over 65% percent identity and an alignment length of at least 25 amino acids. Any reads with high-quality alignments to multiple proteins in the same database were excluded to remove ambiguity in functional gene assignments. Additionally, 1 read had a high-quality alignment to both *mcrA* and the A2-type HCO. This read was also excluded from marker gene analysis.

2.6 Statistics

2.6.1 Summarizing Soil Properties & GHG Fluxes

Differences in soil properties (GWC, pH, NO_3^-, and NH_4^+) and GHG fluxes (CO_2, CH_4, and N_2O from chamber measurements, N_2O from soil-only and soil + NO_3^- + glucose DEA experiments) across ecosystem types were investigated. First, a linear mixed-effects model was fit for each soil property, GHG flux measurement, and DEA flux measurement with the lmer function in R. 'Site' was included as a random variable for all models, 'campaign' was included as a ran-

dom variable for chamber flux measurements, and 'replicate' was included as a random variable for the DEA soil + NO_3^- + glucose treatment. Tukey HSD tests were then performed using the glht function in R to determine the significance of pairwise differences in variables between ecosystem types. Mean values for each variable grouped by ecosystem type are reported in Table 1. The CH_4 Linear model fit and Tukey HSD tests were performed with the one datapoint from an exceedingly high CH_4 flux plot (>3 standard deviations from the mean CH_4 flux) removed. When this outlier plot was included, distinctions between all other CH_4 fluxes were negligible and there were no significant differences between ecosystem types ($p > 0.05$). The removal of this outlier is noted in the Table 1 legend. Appendix A Figures 3 and 4 show boxplots of CH_4 fluxes by ecosystem type with and without the outlier sample.

2.6.2 Determining Microbial Community Structure

General similarities and differences in microbial community genetic structure across ecosystem types were investigated with an ordination of orthology RPKM values. Principal components from the matrix of orthology RPKM values by sample were computed using the prcomp function in R with variable scaling enabled so that all variables had unit variance prior to principle components analysis. The first two principal component vectors were used to investigate microbial community structure (Figure 1). Since only a low percentage of reads were assigned a function based on FMAP, a kmer-based dissimilarity measure was additionally used as a confirmatory comparison of overall genetic composition of samples (MASH, [41], Appendix A Figure 1).

2.6.3 Comparing Marker Gene Abundances Across Ecosystem Types

Marker gene abundance was compared across ecosystem types using a quasi-Poisson model in which gene abundances (count data) were predicted by ecosystem type. Total read depth for each sample was included as an offset in the model to account for varying sampling effort (i.e. sequencing depth). Unlike a standard Poisson model, a quasi-Poisson model allows the model dispersion parameter to vary, which was useful in accounting for overdispersion in the distribution of marker gene abundance data. Model fit was generally good, with the exception of the low-abundance *nirS* and *cNor* genes, for which residuals plots were highly skewed due to a large proportion of zero counts. To highlight that model fit was not ideal for these genes, model coefficient significance ($p < 0.05$) is noted in the figure (Figure 2). The emmeans package in R was used to extract mean gene abundance values (per million total reads) from the fitted models. These values are presented with error bars indicating a 95% confidence interval for the estimate of the mean in Figure 2. Abundances of the denitrification marker genes *nirK*, *nirS*, *cNor*, *qNor*, and *nosZ* were aggregated (Figure 2c) because these genes are unique to the denitrification pathway.

Associations between marker gene abundance and soil properties (pH, soil moisture (GWC), NO_3^-, and NH_4^+) were statistically tested using mixed-effects linear models. Functional gene abundances were first normalized to reads per million total reads for each sample. Next, both the functional gene response variable and all predictor variables were Box-Cox transformed to improve the linearity of the relationship between predictors and response. All variables were then scaled to the standard normal distribution so that a model coefficient rep-

resents the effect of a single standard deviation increase in the Box-Cox transformed independent variable on the Box-Cox transformed dependent variable. This serves to make the coefficients somewhat more interpretable. Ecosystem type was included as a random effect, with site a nested random effect within ecosystem type. Regression coefficients and significance are reported in Table 3. It should be noted that Poisson regressions were also fit to functional gene count data, with total read depth as an offset and an observation-level random effect added to account for overdispersion in the Poisson distribution fit. Poisson regressions yielded similar results for the HCO and methanogenesis marker genes, however for total denitrification gene abundance no associations were significant using Poisson regression. The AIC scores were on the order of 5-10 times greater for linear models than Poisson models, so only linear modeling results are reported.

2.6.4 Relating Marker Gene Abundances and GHG Emissions

Relationships between functional gene abundance and GHG fluxes from chamber flux measurements were investigated using mixed-effects linear modeling. The best marker gene feature set was determined by visual inspection of associations between marker gene abundances and GHG fluxes. Regression models were fit to data from plots for which at least one valid chamber flux measurement was available (n = 31). CO_2 flux was predicted by A1-type HCO gene abundance, CH_4 flux was predicted by *mcrA* gene abundance, and N_2O flux was predicted by total (*nirK* + *nirS* + *cNor* + *qNor* + *nosZ*) denitrification gene abundance. Because both CH_4 fluxes and scaled *mcrA* gene abundance exhibited highly skewed distributions, these variables were log-transformed before

regression, a distinction that is noted in the figure axes. Log-transformation was performed in order to satisfy the assumptions of linear modeling, namely that model residuals are normally distributed with constant variance. All regression models included marker gene abundance as a fixed effect and ecosystem type as a random effect with site as a random effect nested within ecosystem type. Results are presented in Figure 3.

2.6.5 Controls on Denitrification

Motivated by unexpected results from DEA experiments, structural equation modeling (SEM) was used to test a conceptual model of associations between soil properties, including NO_3^- substrate concentration, total denitrification gene abundance, and N_2O flux from the soil-only DEA treatment. Structural equation modeling is a statistical technique useful for testing a conceptual model of linear relationships between multiple variables using observed data. Since no latent variables are included in the model, the problem is equivalent to a path analysis. An important distinction in the model arises between exogenous variables, which are assumed to be independent variables unaffected by any other variables in the model, and endogenous variables, which are assumed to be variables which depend on exogenous and/or other endogenous variables [58, 30]. For the purposes of the moderation hypothesis tested in this study, NO_3^-, NH_4^+, soil moisture (GWC), and pH were assumed to be exogenous variables with direct effects on both denitrification gene abundance and N_2O fluxes. Total (*nirK* + *nirS* + *cNor* + *qNor* + *nosZ*) denitrification gene abundance and N_2O fluxes were endogenous variables in the model. A direct effect between total denitrification gene abundance and N_2O flux was included to test

the moderating impact of denitrification gene abundance on N_2O flux. Since SEM requires continuous variables, total denitrification gene abundance was normalized to the total read depth in each sample so that this variable represents frequencies rather than counts. Total gene abundance, rather than abundance of each individual gene, was included because gene abundances were either highly correlated with one another (e.g. *nirK* and *qNor*) or exhibited zero-inflated distributions that did not satisfy the linear modeling assumptions of the SEM (e.g. *cNor*, *nirS*, *nosZ*). Collapsing denitrification gene abundances additionally served to increase available degrees of freedom for estimating model parameters. All variables except N_2O production, which was log-transformed, were Box-Cox transformed to improve the linearity of relationships, an additional SEM assumption. One outlier DEA measurement had a particularly large negative N_2O flux that was greatly exaggerated by the data transformation. Relationships between all transformed variables both with and without this outlier are reported in Appendix A Figures 14 and 15. The model was fit both with and without this outlier and results from both fits are reported. Variables were then scaled to the standard normal distribution for comparison purposes. The SEM model was fit in STATA with ecosystem type and site collapsed into a single random effect because STATA only allows for two factor levels. The conceptual model was fitted to soil property data, total denitrification gene abundance, and soil-only DEA N_2O flux data from 66 plots (67 with the outlier plot included).

CHAPTER 3

RESULTS

3.1 Ecosystem Type Characterizations

Soil properties differed significantly depending on ecosystem type (Table 1). Emergent wetland and scrub-shrub wetland soils were highly saturated, having the highest moisture contents of soils from any ecosystem type. Soil pH was generally acidic across all ecosystem types, ranging from 4.1 in upland soils to 5.3 in riparian soils with salmon. Soil NO_3^- was uniformly low except in riparian soils with salmon, where it was on the order of 10 times higher than in other soils. Soil NH_4^+ was highest in upland and forested wetland soils (88.4 and 57.5 μg g soil^{-1}, respectively) and lowest in emergent wetland soils (18.6 μg g soil^{-1}). Soil phosphorus (P) was negligible in the emergent and scrub-shrub wetland soils as well as in the riparian soils with and without salmon. Soil P was highest in upland and forested wetland soils (18.3 and 10.2 g μg soil^{-1}, respectively).

Table 1: Average measures of soil characteristics and GHG fluxes determined via the static chamber flux method in each ecosystem type. Letter columns represent statistically significant differences between ecosystem types ($p < 0.05$). *A single exceedingly high CH_4 flux value was removed from the dataset prior to ANOVA and Tukey HSD tests for CH_4.

Ecosystem Type	GWC (%)		pH		NO_3^- (μg g soil^{-1})		NH_4^+ (μg g soil^{-1})		P (μg g soil^{-1})		CO_2 Chamber Flux (mg CO_2-C m^{-2} hr^{-1})		CH_4 Chamber Flux* (μg CH_4-C m^{-2} hr^{-1})		N_2O Chamber Flux (μg N_2O-N m^{-2} hr^{-1})	
U	3.82	a	4.08	a	3.19	a	88.37	b	18.31	a	157.18	b	-14.36	a	19.58	a
PFO	4.99	a	4.54	ac	5.22	a	57.48	ab	10.16	a	159.06	b	640.28	ab	-8.62	a
PSS	17.70	b	4.21	ab	4.81	a	19.67	a	2.25	a	47.63	a	3553.83	b	-7.21	a
PEM	11.21	ab	4.56	ac	5.19	a	18.61	a	0.41	a	46.07	a	2244.81	ab	29.66	a
RNF	1.37	a	5.08	bc	2.65	a	23.07	a	0.00	a	154.03	ab	-80.01	a	11.06	a
RF	0.97	a	5.30	c	33.12	a	35.98	a	0.08	a	108.67	ab	47.91	a	23.70	a

Table 2: Average DEA N_2O-N fluxes in each ecosystem type, as well as the percent change in average N_2O flux measured between soil-only and soil + NO_3^- + glucose treatments. Letter columns represent statistically significant differences between ecosystem types ($p < 0.05$).

Ecosystem Type	Soil-Only DEA N_2O Flux (μg N_2O-N g soil^{-1} day^{-1})		Soil + NO_3^- + Glucose DEA N_2O Flux (μg N_2O-N g soil^{-1} day^{-1})		% Increase in N_2O Flux
U	0.1	b	17.4	ab	14722.5
PFO	0.3	b	34.6	a	10917.9
PSS	1.6	b	14.5	ab	782.5
PEM	0.5	b	15.3	ab	2834.5
RNF	0.8	b	3.3	b	302.5
RF	5.3	a	10.8	b	102.5

3.2 GHG Chamber Flux Measurements

To measure soil GHG production from the three respiratory pathways considered, the gaseous products CO_2, CH_4, and N_2O were quantified from soils in each ecosystem type. The flux of each gas from soils was first measured in situ using the static chamber flux method (see 'Chamber Flux'measurements in Table 1). CO_2 and CH_4 production showed opposite trends across ecosystem types, with the emergent and scrub-shrub wetland soils producing the least CO_2 and the most CH_4. CH_4 production in the upland, forested wetland, and riparian soils with and without salmon was on the order of to 10 to 100 times smaller than in the emergent and scrub-shrub wetland soils. In contrast, CO_2 production in upland, forested wetland, and riparian soils with and without salmon soils was on the order of 10 times greater than in the emergent and scrub-shrub wetland soils. In contrast, N_2O was barely quantifiable via the static chamber flux method in any soils, ranging from a low of -10.1 μg N_2O-N m^{-2} hr^{-1} in forested wetland soils to a high of 44.2 μg N_2O-N m^{-2} hr^{-1} in emergent wetland soils.

3.3 Denitrification Enzyme Assays

DEA experiments were also performed to quantify potential denitrification rates (Table 1). In the soil-only DEA treatment, N_2O production was still universally low in most soils (0.1-1.6 μg N_2O-N m^{-2} hr^{-1}), with the notable exception of soils from riparian sites with salmon. Soil samples from riparian sites with salmon produced on average 5.3 μg N_2O-N m^{-2} hr^{-1}, 3 to 50 times higher than soil samples from other ecosystem types. However, results differed from

the soil-only treatment to the soil + NO_3^- + glucose treatment. In the presence of added NO_3^- and glucose, N_2O production increased between approximately 100% for soils from riparian sites with salmon and 15000% for upland soils. Forested wetland soils produced the most N_2O under the soil + NO_3^- + glucose treatment, with riparian soils with and without salmon producing the least N_2O.

3.4 Microbial Community Composition

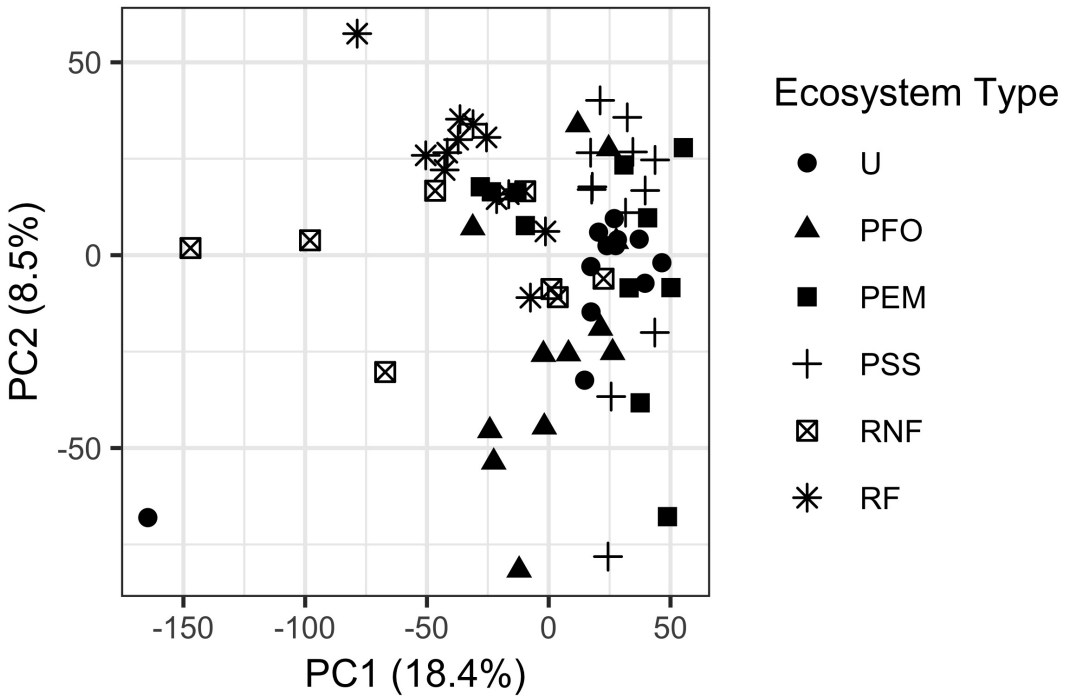

Figure 1: Ordination of study sites by ecosystem type. Principle components are calculated based on KEGG orthology abundance at each site.

An ordination was performed to investigate the microbial community ge-

netic structure across ecosystem types (Figure 1). The ordination is based on the first two principle components calculated from the matrix of orthology abundances across samples determined by FMAP [29]. Together, the first two principle components explain 27% of the variation in the dataset. The only apparent distinctions in community composition are between the riparian sites and the wetland/upland sites. Dissimilarity-based principal coordinates ordination of sample kmer contents (Appendex A Figure 1) shows an even more pronounced separation of riparian sites from sites of other ecosystem types.

3.5 Respiratory Pathway Marker Genes

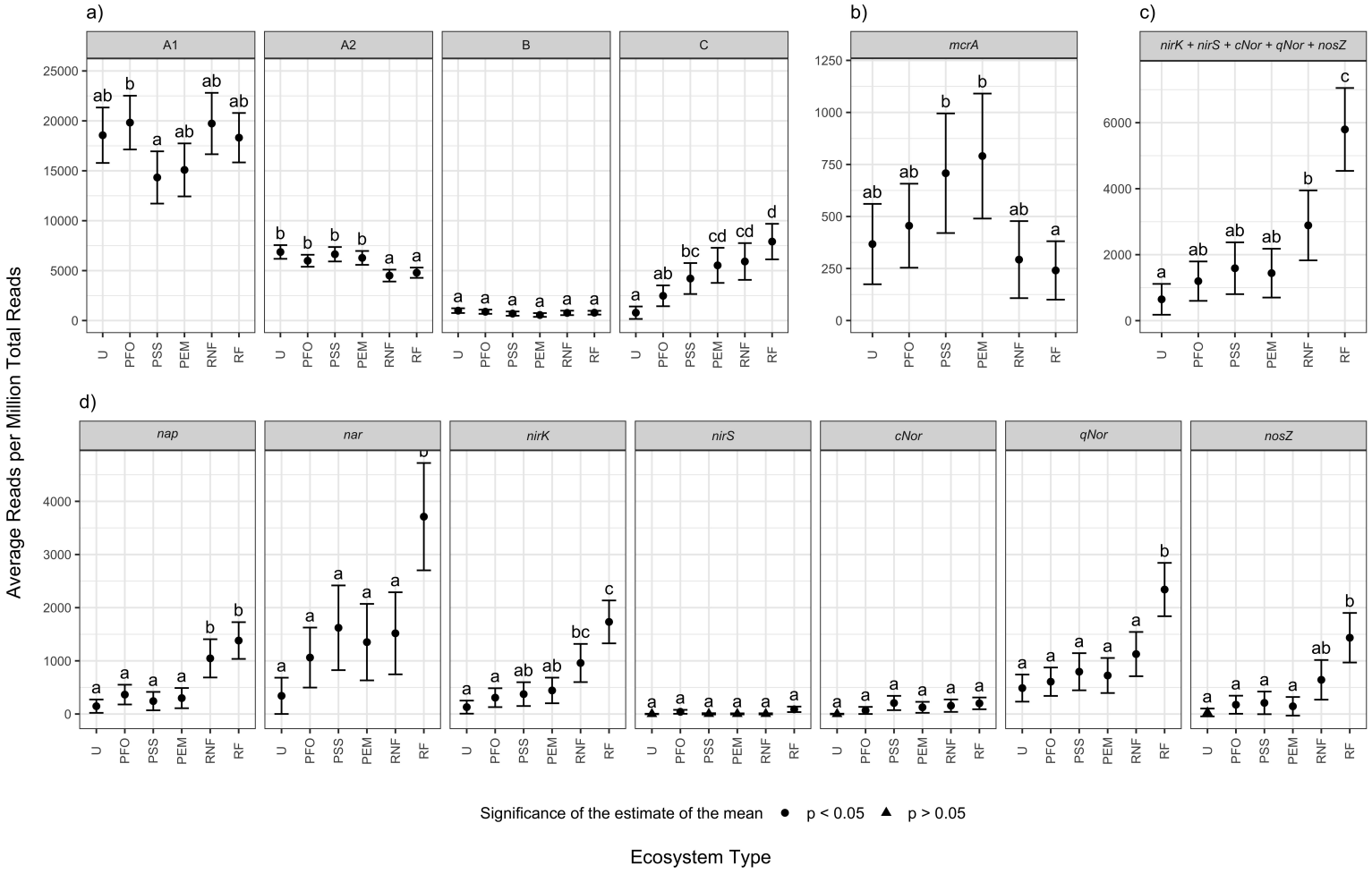

Figure 2: Mean marker gene abundance by ecosystem type. Error bars indicate a 95% confidence interval for estimates of mean abundance. The shape of datapoints indicates whether the quasi-Poisson model coefficient for the ecosystem type was significant since some models were poor fits to zero-inflated marker gene distributions.

3.5.1 Heme-Copper Oxidase Gene Type Abundance

The low-O_2 affinity A1 and A2 -type HCOs were generally more abundant than B and C-type HCOs across ecosystem types. Low O_2-affinity A1-type HCOs were most abundant at forested wetland sites and were significantly less abundant at scrub-shrub wetland sites. High-O_2 affinity C-type HCOs showed a more distributed trend in abundance across ecosystem types, with the lowest C-type HCO abundance occurring in upland soils and an increasing abundance in forested wetland, scrub-shrub wetland, emergent wetland, and riparian sites without salmon. C-type HCO abundance was greatest at riparian sites with salmon. In contrast, A2-type HCO abundance was significantly lower at riparian sites than any other ecostsytem type. Finally, B-type HCO abundance was uniformly low and not significantly different between ecosystem types.

3.5.2 Methanogenesis Gene Abundance

The *mcrA* marker gene for methanogenesis was found in soils from all ecosystem types. *mcrA* abundance was, on average, highest in soils from the scrub-shrub and emergent wetland ecosystem types. However, the only statistically significant differences in *mcrA* abundance were between the highest *mcrA* counts from scrub-shrub and emergent wetland soils and the lowest *mcrA* counts in riparian soils with salmon.

3.5.3 Denitrification Gene Abundance

The magnitude of reads from denitrification genes varies drastically among genes, with *nar* the most abundant gene, on average, and *nirS* the least abundant gene, on average. No significant patterns of variance among ecosystem types are apparent in the low-abundance *nirS* and *cNor* genes. For the more abundant *nap*, *nar*, *nirK*, *qNor*, and *nosZ* genes, a somewhat consistent trend in abundance across ecosystem types is apparent, with lowest abundance in upland soils and highest abundance in riparian soils with salmon. Total denitrification gene abundance was least in upland soils and increased in abundance from forested wetland to emergent wetland, scrub-shrub wetland, and riparian soils without salmon. Total denitrification gene abundance was greatest in riparian soils with fish.

3.6 Relating Marker Gene Abundances and Soil Properties

To investigate the association between measured soil properties and marker gene abundance, mixed-effects linear modeling was employed. Model coefficients are reported in Table 3. It must be noted that coefficients represent the association between Box-Cox transformed, scaled variables. Therefore, it is not possible to easily interpret coefficient magnitudes. Instead, the significance and direction of association are more important to note. Total (*nirK* + *nirS* + *cNor* + *qNor* + *nosZ*) denitrification gene abundance was most significantly positively associated with pH, and was negatively associated with soil moisture (GWC). *mcrA* abundance was positively associated with soil moisture, while A1-type HCO abundance was weakly negatively associated with soil moisture. B-type

Table 3: Table 3. Mixed-effects linear model coefficients for measured soil properties. Asterisks indicate statistical significance of coefficient values (***p < 0.001, **p < 0.01, *p < 0.05, .p < 0.1). Coefficients are between Box-Cox transformed, scaled variables.

Regression Coefficients	Total Denitrification Genes	*mcrA*	A1 HCO	A2 HCO	B HCO	C HCO
pH	0.41***	0.08	-0.28	-0.01	-0.33 .	0.45***
GWC	-0.35*	0.56**	-0.38 .	0.30	-0.33	0.04
NO_3^-	0.01	0.03	-0.001	0.02	0.07	0.11
NH_4^+	-0.14	0.03	0.06	0.14	-0.05	-0.17*

HCO abundance was weakly negatively associated with pH. C-type HCO abundance was strongly positively associated with pH and negatively associated with NH_4^+.

3.7 Relating Marker Gene Abundances and GHG Emissions

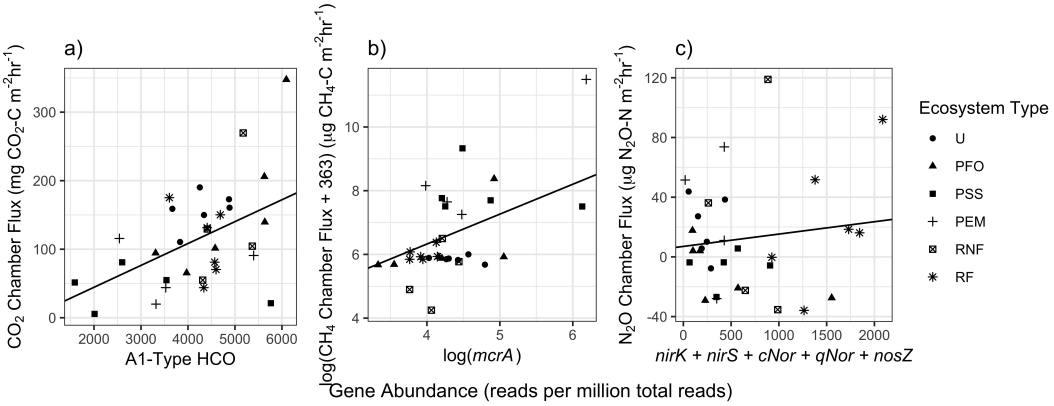

Figure 3: Linear mixed-effects regression of GHG emissions by marker gene abundance based on sample plots with valid chamber flux measurements (n = 31). Shape indicates ecosystem type, and line represents mixed-effects regression fit.

Marker gene abundances were more closely associated with *in-situ* chamber measurements of CO_2 and CH_4 flux than N_2O flux (Figure 3). Regression coefficients for the fixed-effect marker gene variable were significant for CO_2 ($p < 0.01$) and CH_4 ($p < 0.01$), but not for N_2O ($p > 0.05$).

3.8 Controls on Denitrification

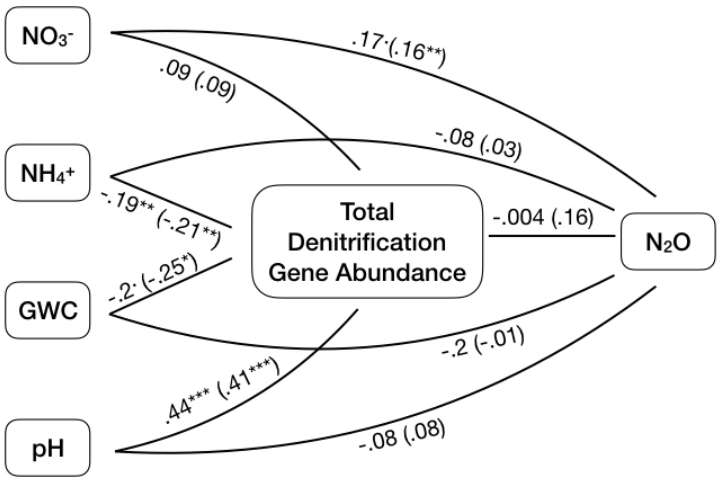

Figure 4: Fitted SEM pathway coefficients. Significance is indicated by asterisks (***p < 0.001, **p < 0.01, *p < 0.05, .p < 0.1). Values outside parentheses are model coefficients from the model fit without the outlier (n = 66) and values inside parentheses are model coefficients from the model fit with the outlier (n = 67).

Structural equation modeling identified significant direct effects between soil properties, total (*nirK* + *nirS* + *cNor* + *qNor* + *nosZ*) denitrification gene abundance, and denitrification activity as measured by DEA N_2O production from the soil-only treatment. No significant indirect effects between soil properties, denitrification gene abundance, and denitrification activity were identified. Higher pH, lower soil moisture, and lower NH_4^+ concentration were associated with higher denitrification gene abundance. Higher NO_3^- was associated with greater denitrification activity. Due to the inclusion of random effects in the model, a p-score is not available. The model AIC score was 197.

CHAPTER 4

DISCUSSION

Study results show that landscape-scale patterns in respiratory functional gene abundance were evident and were associated with gradients in soil moisture (GWC), pH, NO_3^-, and NH_4^+ concentration across PCTR ecosystem types (Table 3). Measured GHG emissions also varied significantly by ecosystem type, with the greatest CO_2 emissions produced in upland and forested wetland ecosystems and the greatest CH_4 emissions produced in emergent and scrub-shrub wetland ecosystems. In contrast, observed N_2O emissions were negligible and were not significantly different across ecosystem types (Table 1). *In-situ* measured CO_2 emissions were positively associated with A1-type HCO gene abundance and CH_4 emissions were positively associated with *mcrA* gene abundance (Figure 3a,b). However, *in-situ* measured N_2O emissions were not associated with total (*nirK* + *nirS* + *cNor* + *qNor* + *nosZ*) denitrification gene abundance (Figure 3c). While measured N_2O emissions were negligible across ecosystem types, denitrification marker gene abundance and soil-only DEA experiments suggest that the greatest denitrification activity likely occurs at riparian sites with salmon (Figure 2c, Table 1). Despite providing marker gene evidence of microbial community specialization for denitrification in riparian soils with salmon, soil NO_3^- was more strongly associated with denitrification activity and N_2O emissions in the DEA experiments than denitrification gene abundance (Figure 4). These results demonstrate that microbial community genetic structure may be a useful indicator of the type and magnitude of GHG emissions from various PCTR ecosystem types, though in the case of denitrification and N_2O at least, substrate concentration is, as expected in an N-limited environment, a stronger controller of N_2O emissions than community genetic capacity.

Should N-deposition increase in the future due to increased fossil fuel combustion or altered weather patterns, this study indicates PCTR microbial communities are poised to respond with increased denitrification activity across ecosystem types, potentially increasing the N_2O contribution of a wide swath of PCTR soils.

4.1 Variability in Respiratory Marker Gene Abundance by Ecosystem Type

Marker gene abundance varied in expected ways across ecosystem types for each respiratory pathway considered. These distinctions were independent of larger trends in overall community genetic composition, in which only riparian soils stand out from other ecosystem types (Figure 1). Furthermore, patterns in marker gene abundance between sites were strongly associated with various soil properties for each respiratory pathway (Table 3). These results provide evidence that the topographically-driven hydrologic differences among PCTR ecosystem types, which lead to the formation of distinct plant communities, also drive respiratory specialization in below-ground microbial communities.

An inverse relationship was observed between A1-type heme-copper oxidase gene abundance and *mcrA* gene abundance across ecosystem types (Figure 2a and 2b), a relationship that is consistent with the differing ecological niches fulfilled by organisms carrying these genes. It is expected that methanogens would have a selective advantage over aerobic microorganisms in the nearly constantly saturated scrub-shrub and emergent wetland soils, while specialized aerobes carrying the low O_2-affinity A1-type HCO would fare better in soils

where the water table is typically lower, such as upland and forested wetland sites. In support of this conclusion, both *mcrA* and A1-type HCO gene abundance were found to be most closely associated with soil moisture out of all measured soil properties (Table 1). As expected, CH_4 emissions were significantly greater from emergent and scrub-shrub wetland sites than other sites, while CO_2 emissions were significantly greater at upland and forested wetland sites than emergent and scrub-shrub wetland sites (Table 1). CO_2 emissions from riparian sites with and without salmon were intermediate (Table 1).

Interestingly, the high O_2-affinity C-type heme-copper oxidase gene was more closely associated with pH and NH_4^+ concentration than soil moisture (Table 3). In general, abundance of denitrification genes followed the same trend across ecosystem types as the C-type HCO gene, suggesting a potential regulatory or functional linkage between the C-type HCO gene and denitrification genes (Figure 2a,d). One possible explanation for higher C-type and lower A2-type HCO abundance at riparian sites without salmon is taxonomic overlap with denitrifiers. According to taxonomic assignment of HCO and denitrification gene reads (Appendix A, Figure 13), C-type HCO reads were overwhelmingly from proteobacteria (91% of taxonomically classified C-type HCO reads), while only about half of taxonomically classified A2-type HCO reads were from proteobacteria. Denitrification marker genes were most commonly found among proteobacteria as well (83% of taxonomically classified denitrification marker gene reads), suggesting that proteobacteria which contain denitrification genes may also preferentially have C-type HCOs in contrast to the more taxonomically diverse A2-type HCOs. Additionally, the C-type HCO may serve to detoxify nitric oxide (NO), a cytotoxic compound, at sites with higher NO-producing denitrification activity [44]. Finally, the C-type HCO is known

to serve a regulatory function in response to anoxic conditions, which also promote denitrification [39]. While we can only speculate as to the nature of any regulatory or functional linkage between C-type HCO and denitrification genes, our results suggest that such a linkage may occur in PCTR soils.

Variability in the distribution of individual denitrification genes across ecosystem types serves as a further indication of microbial respiratory specialization to unique soil conditions across PCTR ecosystem types. Of the genes unique to the denitrification pathway, *qNor*, *nirK*, and *nosZ* were the most abundant. *nirS* and *cNor* abundances were very low (on the order of 10 reads/million total reads) across all ecosystem types. The *nirS* and *cNor* enzymes each require additional proteins for enzyme assembly [49], making these genes less energetically favorable than their functionally redundant counterparts *nirK* and *qNor*. *nirS* and *cNor* abundance may be an indicator of conditions highly favorable for denitrification (Roco & Nadeau, in preparation). Therefore, the uniformly low NO_3^- levels across ecosystem types in the PCTR, even at riparian sites with salmon where NO_3^- is appreciably greater than at other sites, may not be sufficient to support specialist denitrifier communities. Even the higher NO_3^- concentrations found in riparian soils with salmon may be ephemeral, since measurements were made during the time salmon were in the stream, which occurs for only a limited portion of the year. The denitrification genes most represented in this study, *qNor*, *nirK*, and *nosZ*, do represent a complete denitrification pathway and abundance of each of these genes was generally similar across the upland, wetland, and riparian sites without salmon and highest at riparian sites with salmon. Among the genes not unique to denitrification, *nap* (the periplasmic NO_3^- reductase) abundance was similar between riparian sites with and without salmon. In contrast, *nar* (the cytoplasmic NO_3^- reductase) abundance

was greater only at riparian sites with fish, consistent with abundance patterns of the other highly-abundant denitrification genes (Figure 2d). This distinction could be due to the differing modes of action of these NO_3^- reductases; *nar* is definitively and commonly involved in NO_3^- respiration while the significance of *nap* to NO_3^- respiration is debatable [52]. Greater *nar* abundance at riparian sites with salmon suggests that higher NO_3^- concentrations, evidently from salmon presence, might support microbial communities with greater NO_3^- respiratory capacity in riparian soils adjacent to salmon-accessible streams.

4.2 Relating Marker Gene Abundances and GHG Emissions

This study presents evidence that microbial community respiratory specialization (i.e. variability in marker gene abundance) is linked to summer GHG emissions. *In-situ* measurements of CO_2 and CH_4 emissions from two summertime sampling campaigns using the chamber flux method were significantly associated with A1-type HCO and *mcrA* marker gene abundances, respectively (Figure 3a,b). These results suggest that marker gene abundance may be a useful indicator of CO_2 and CH_4 emissions. However, this conclusion requires further testing, especially considering that our data only represent limited intra-seasonal temporal variability in gas fluxes and no temporal variability in marker gene abundance. Furthermore, N_2O emissions were not associated with total (*nirK* + *nirS* + *cNor* + *qNor* + *nosZ*) denitrification gene abundance (Figure 3c). Either low denitrification activity in general or high activity of the nitrous oxide reductase (NosZ) enzyme may have resulted in the low N_2O emissions observed. According to the "leaky pipe" model proposed by [18], under conditions which stimulate high denitrification activity, high concentrations of N_2O

may overwhelm active sites of NosZ enzymes and be ”leaked” from the denitrification pipeline to the atmosphere. However, under conditions of low denitrification this phenomenon is likely not observable, as may have been the case during our chamber flux sampling campaigns. Denitrification activity is known to be highly temporally variable and is therefore difficult to asses based on chamber sampling measurements [22]. However, several factors suggest that denitrification activity is greater at riparian sites with salmon than other sites, even if this difference did not result in enhanced N_2O emissions during our chamber flux measurement campaigns. Firstly, denitrification gene abundance was significantly higher in riparian soils with salmon than other soils (Figure 2c), indicating that conditions in which denitrification is necessary and N-oxide substrates are available may occur more often in these soils than other PCTR soils. Secondly, greater concentrations of soil NO_3^- were measured in riparian soils with salmon while salmon were in the stream, meaning substrate availability for denitrification is at least temporarily enhanced in these soils (Table 1). Finally, DEA measurements from the soil-only treatment indicate that, unlike in the field, under controlled laboratory conditions (possibly due to higher temperature than in the field), denitrification activity is greater at riparian sites with salmon (Table 2). Each of these pieces of evidence indicates that conditions for denitrification are more favorable at in riparian soils with salmon than other PCTR soils. Following the leaky-pipe model, it is plausible that over longer time-scales than were measured during this study appreciable quantities of N_2O may be ”leaked” from the denitrification pipeline in riparian soils with salmon.

4.3 Controls on Denitrification

In answer to our third study question, DEA results and confirmatory analysis with structural equation modeling suggest that total denitrification gene abundance does not have a significant controlling effect on denitrification activity in PCTR soils, at least under the DEA conditions studied. The contrast between DEA results from the soil-only treatment and the soil + NO_3^- + glucose treatment indicated that either NO_3^- or glucose was a limiting factor in denitrification, rather than denitrification gene abundance. Soils from riparian sites with salmon exhibited the greatest denitrification activity in the soil-only treatment consistent with them having the highest NO_3^- concentrations and the greatest total denitrification gene abundance (Table 1, Figure 2c). However, upon addition of NO_3^- and glucose to PCTR soils, soils with smaller total denitrification gene abundances exhibited the greatest increases in denitrification activity compared to the soil-only DEA experiment (Table 1). Strikingly, riparian soils with and without salmon exhibited the lowest denitrification activity in the soil + NO_3^- + glucose treatment (Table 1). The explanation for this observation is not readily apparent, though it has been noted that glucose is a somewhat selective C source [38]. It is possible that the microbial communities in the non-riparian soils were better able to capitalize on provided N and C resources than riparian soils in the soil + NO_3^- + glucose DEA treatment. Regardless, these results suggest that the total abundance of denitrification genes is not necessarily a controlling factor for denitrification rates, at least under the DEA conditions studied.

To statistically test whether denitrification gene abundance controlled the effect of NO_3^- concentration and other soil properties on denitrification activity, structural equation modelling was used. SEMs have previously been used to

infer causal relationships between soil properties, including substrate concentrations, microbial gene abundance, and process rates in the context of denitrification [45] as well as other C and N cycling processes [32]. In this study, we used SEM to evaluate the hypothesis that denitrification gene abundance moderates the effect of soil properties, including NO_3^- concentration, on denitrification activity as measured by the soil-only DEA treatment. Congruent with linear modeling results (Table 3), pH, soil NH_4^+, and moisture content were significantly associated with total (*nirK* + *nirS* + *cNor* + *qNor* + *nosZ*) denitrification gene abundance according to SEM (Figure 4). However, the only significant association with N_2O production (our proxy for denitrification activity) was NO_3^- concentration. No indirect relationships between soil properties, total denitrification gene abundance, and N_2O emissions were statistically significant ($p > 0.05$). These results are inconsistent with those achieved by Petersen *et al.* ([45]), who concluded that *nirK* and *nosZ* gene abundance, quantified via qPCR, were strongly associated with potential denitrification rates from Southeast Alaskan soils. However, Petersen *et al.* used fertilized DEA experiments including added NO_3^- and glucose for their SEM analysis. A similar analysis performed by Lammel *et al.* ([32]) in Amazonian soils using in-situ chamber-flux measurements found a weaker relationship between functional gene abundance and CO_2, CH_4, and N_2O emissions. Our study provides new information on the relationship between soil properties, including NO_3^- concentration, on denitrification activity in PCTR soils under conditions which mimic field conditions more closely than DEA experiments with added N and C.

4.4 Conclusion

In conclusion, we provide evidence that microbial communities specialize in their respiratory capabilities across PCTR ecosystem types, which has implications for where in the landscape we can expect the greatest emissions of various greenhouse gas to come from. We present evidence that the greatest CO_2, CH_4, and N_2O emissions are partitioned by ecosystem type in the PCTR, with CO_2 emissions highest in upland and forested wetland sites, CH_4 emissions highest in emergent and scrub-shrub wetland sites, and N_2O emissions *potentially* highest in riparian sites with salmon. Respiratory marker gene abundance was a useful indicator of these landscape-scale distinctions in GHG emissions for at least CO_2 and CH_4. Based on denitrification activity measurements, we hypothesize that marker gene abundance is also predictive of longer-term N_2O emissions, though we provide no direct evidence for this. Additionally, DEA experiments and confirmatory structural equation modeling suggest that NO_3^- concentration is a stronger control on denitrification activity than denitrification gene abundance in these soils. This result implies that should N deposition increase in the PCTR, we can expect increased N_2O emissions from soils across all ecosystem types.

APPENDIX A

SUPPLEMENTAL FIGURES

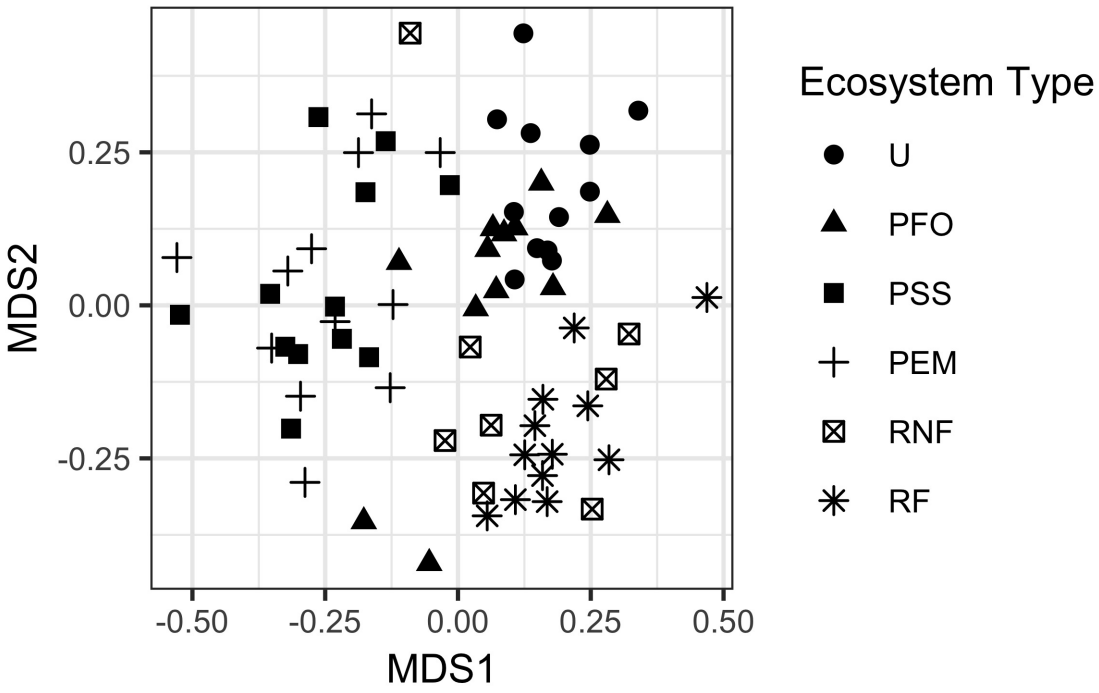

Figure 1: Non-metric dimensional scaling (NMDS) based on dissimilarity measures between read sets from each ecosystem type. Dissimilarity was calculated using MASH [41] with default settings. NMDS coordinates were determined using the metaMDS function in R.

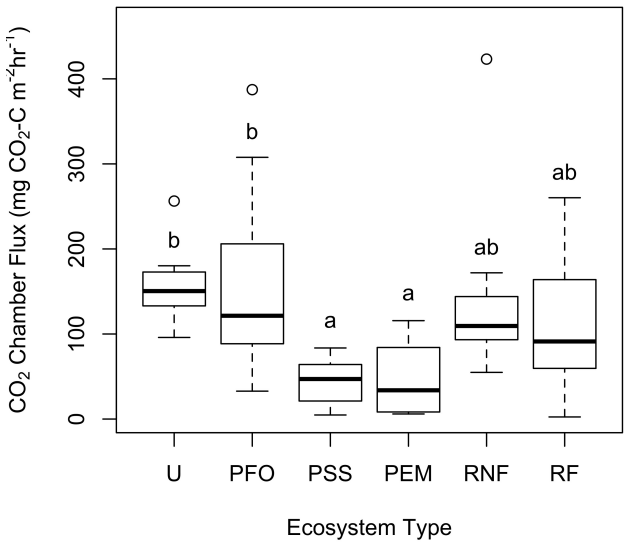

Figure 2: Chamber CO_2 flux measurements by ecosystem type.

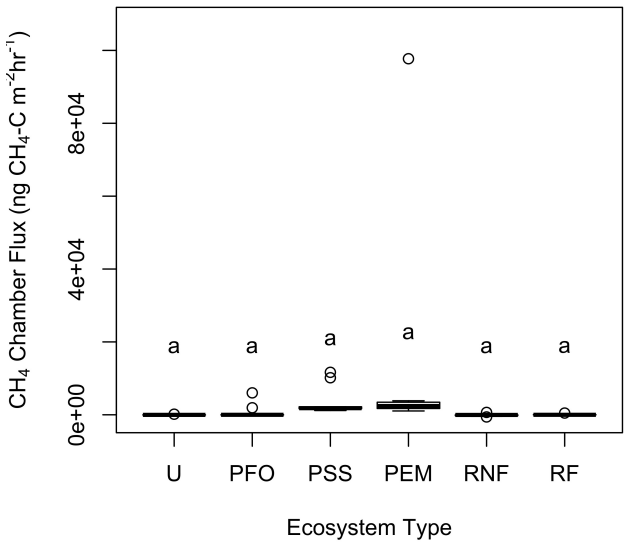

Figure 3: Chamber CH_4 flux measurements by ecosystem type, with outlier plot included.

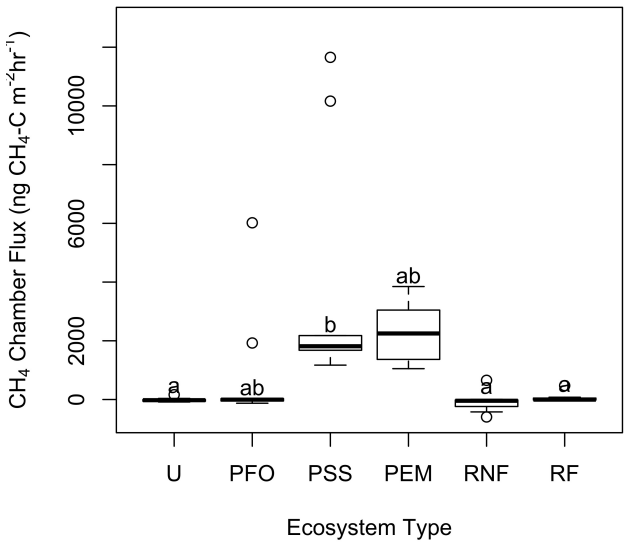

Figure 4: Chamber CH_4 flux measurements by ecosystem type, with outlier plot removed.

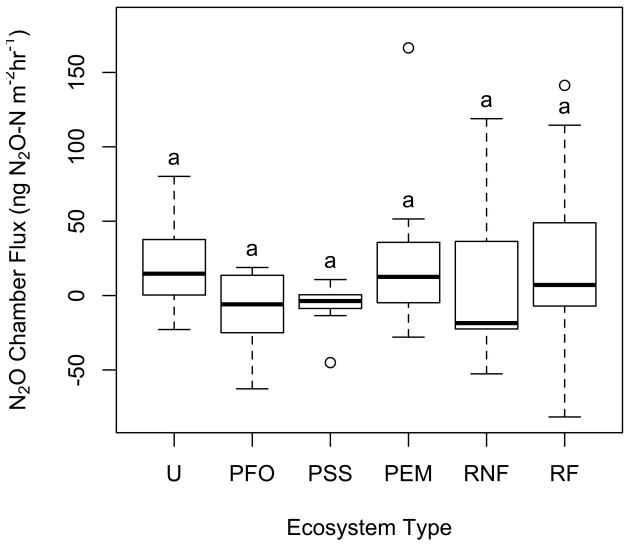

Figure 5: Chamber N_2O flux measurements by ecosystem type.

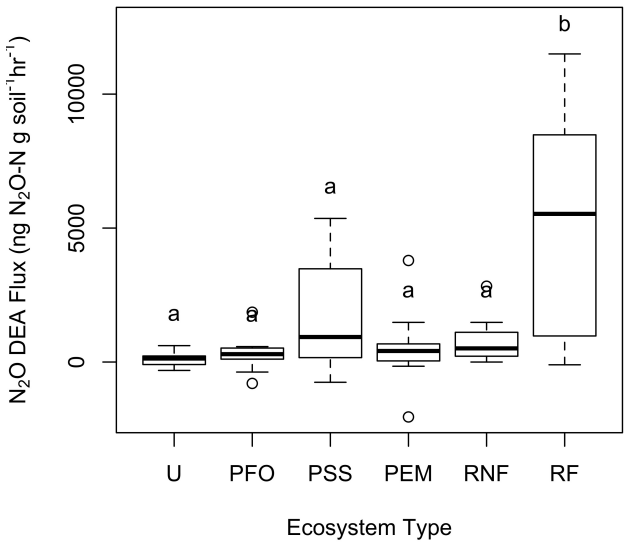

Figure 6: DEA N_2O flux measurements by ecosystem type (soil-only treatment).

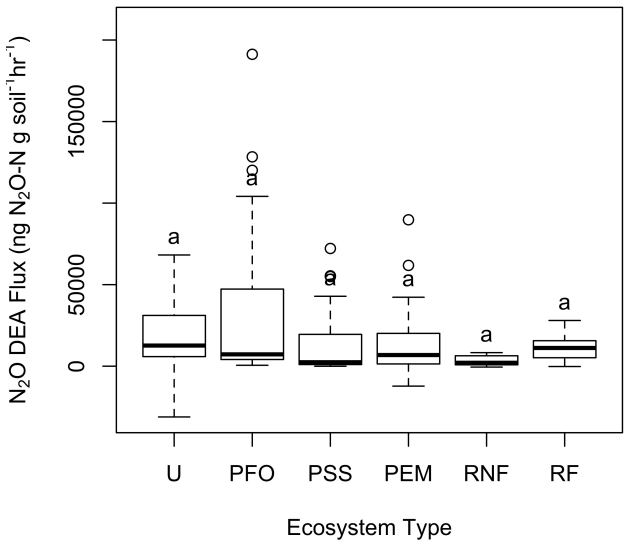

Figure 7: DEA N_2O flux measurements by ecosystem type (Soil + NO_3^- + glucose treatment).

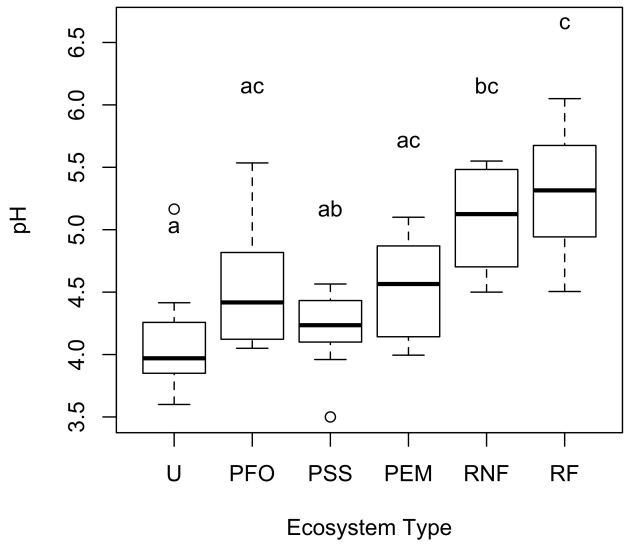

Figure 8: pH measurements by ecosystem type.

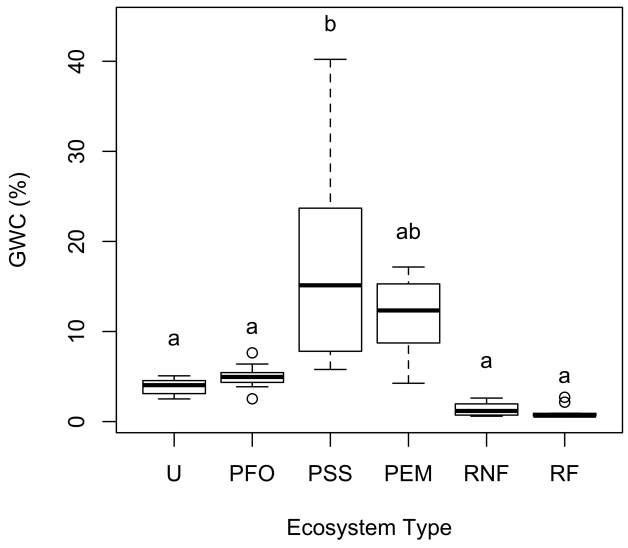

Figure 9: Gravimetric water content measurements by ecosystem type.

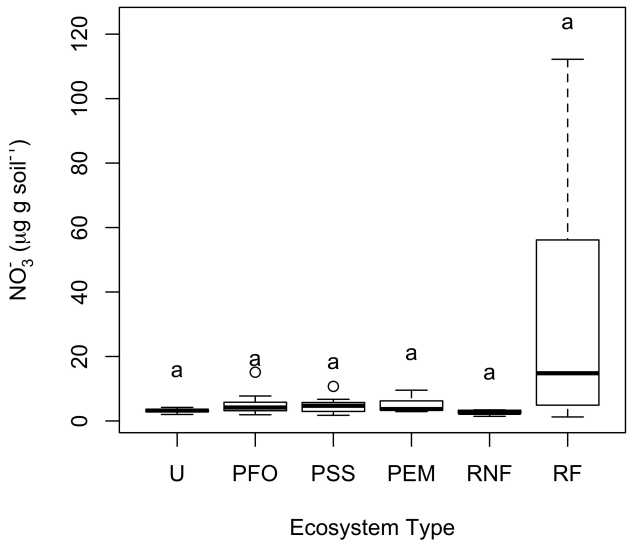

Figure 10: NO_3^- measurements by ecosystem type.

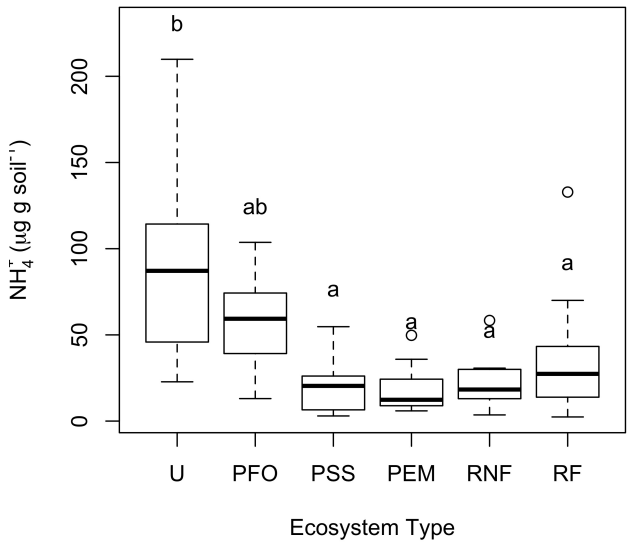

Figure 11: NH_4^+ measurements by ecosystem type.

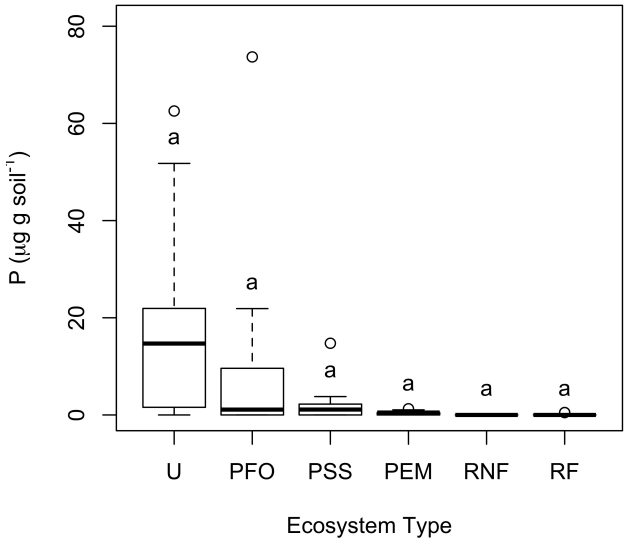

Figure 12: P measurements by ecosystem type.

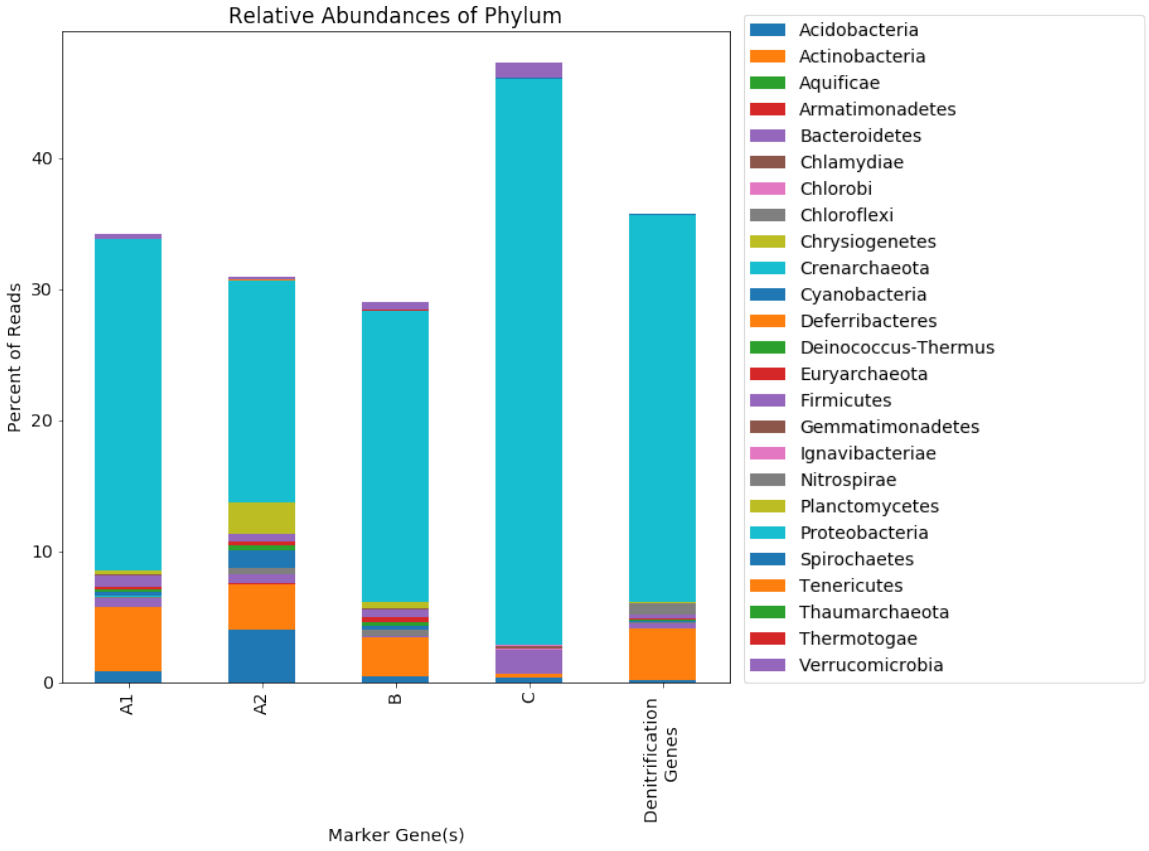

Figure 13: Taxonomic assignment of heme-copper oxidase and denitrification marker gene reads aggregated at the phylum level. Taxonomy assignment was performed using Centrifuge [28] with default settings.

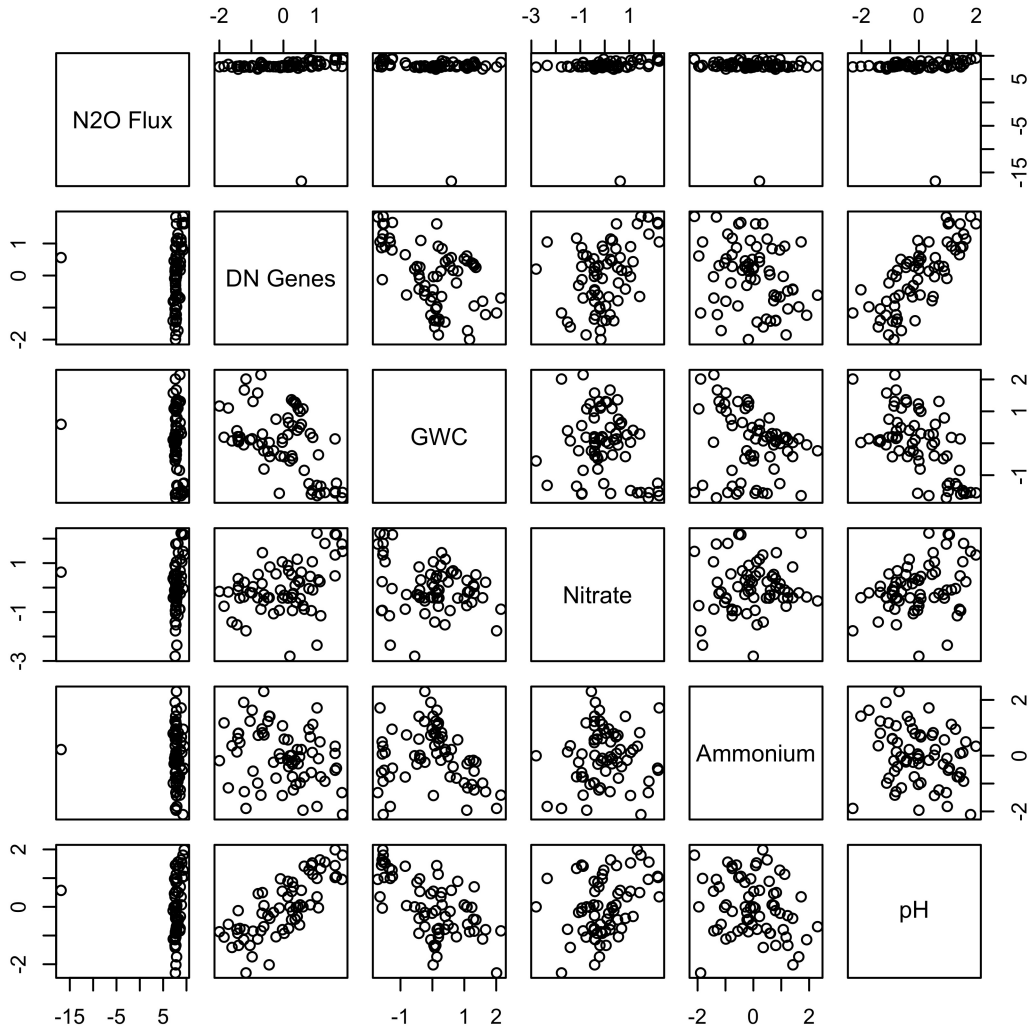

Figure 14: Associations between SEM variables with outlier (n = 67).

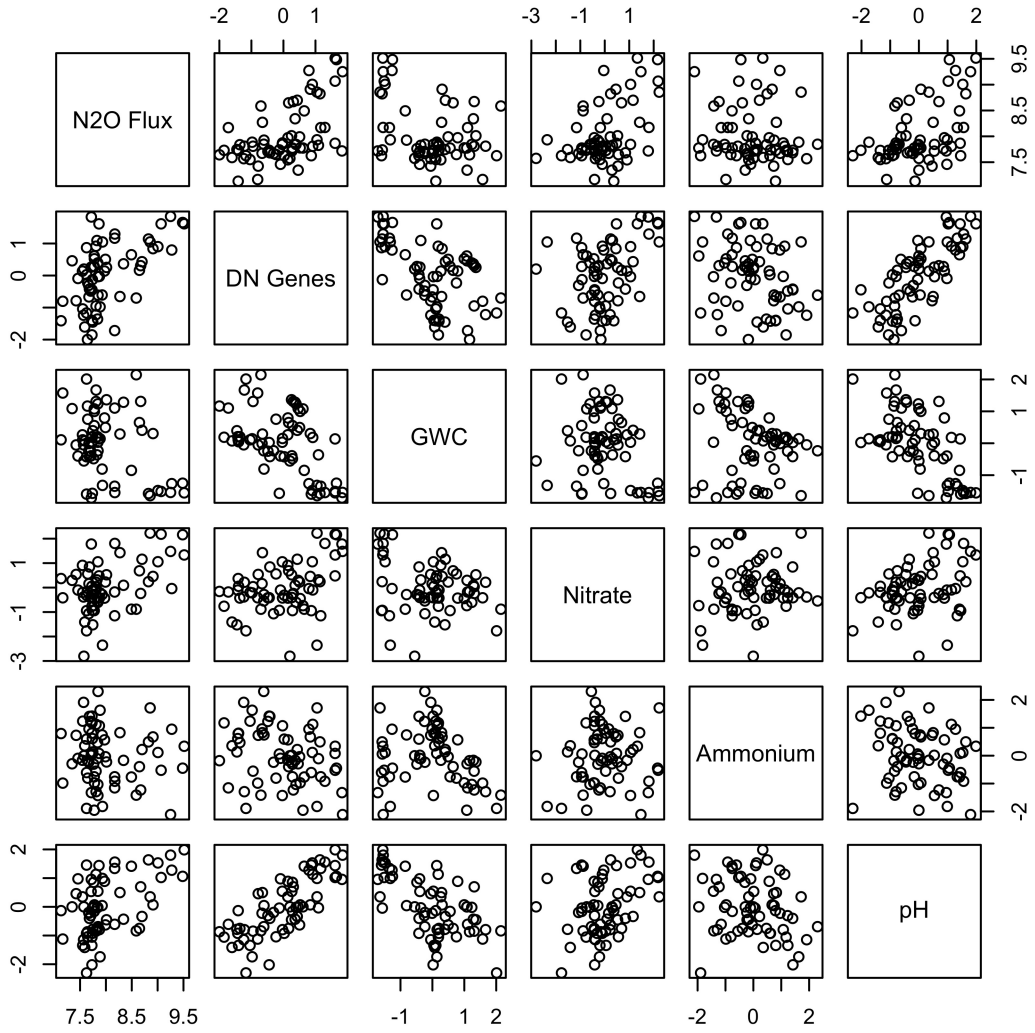

Figure 15: Associations between SEM variables without outlier (n = 66).

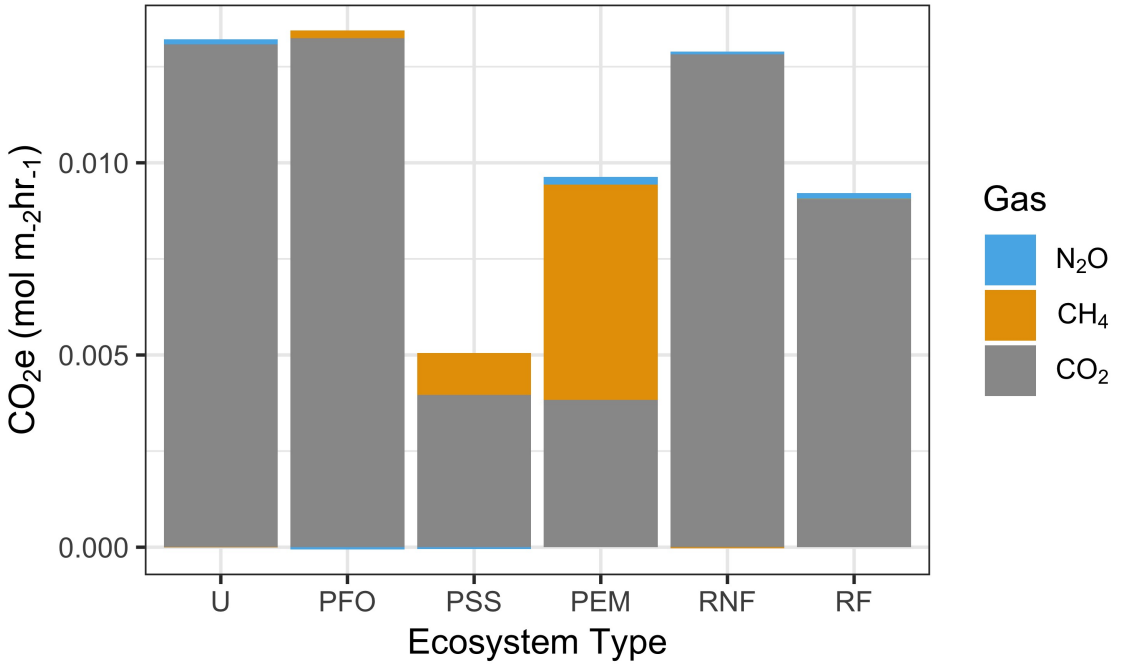

Figure 16: Average GHG emissions from each ecosystem type (measured by chamber flux method) by gas in CO_2 equivalents (CO_2e)

APPENDIX B

SUPPLEMENTAL TABLES

Table 1: Protein database sequence composition.

Gene Type	Sequence Count
A1	947
A2	143
B	116
C	365
nap	1976
nar	3066
nirK	660
nirS	205
cNor	530
qNor	592
nosZ	484
mcrA	1460

Table 2: Metagenomic read counts.

Ecosystem Type	Site	Plot	Sample read depth Before QC	Sample read depth After QC	Average read length	Total reads assigned FMAP	% Reads assigned FMAP
RF	13	A	4681482	4681482	141.23	328237	7.01
RF	13	B	4726447	4726445	142.01	331075	7
RF	13	C	4431510	4431508	142.53	316826	7.15
RF	13	D	3592139	3592139	139.5	246182	6.85
RF	16	A	4904794	4904793	141	340040	6.93
RF	16	B	5324731	5324731	140.07	342251	6.43
RF	16	C	4577941	4577941	139.96	298259	6.52
RF	16	D	4443496	4443495	141.67	267954	6.03
RF	17	A	4733232	4733225	140.23	304804	6.44
RF	17	B	6180721	6180721	141.16	434519	7.03
RF	17	C	4988502	4988500	142.41	365641	7.33
RF	17	D	4970928	4970927	141.17	296346	5.96
RNF	14	A	3648687	3648685	142.89	239452	6.56

RNF	14	B	2881419	2881419	140.97	189470	6.58
RNF	14	C	3624674	3624672	142.8	256896	7.09
RNF	14	D	5037842	5037837	142.75	356489	7.08
RNF	15	A	7132422	7132418	137.46	471944	6.62
RNF	15	B	8082925	8082925	137.15	567582	7.02
RNF	15	C	5493437	5493431	141.19	388238	7.07
RNF	15	D	4347076	4347072	140.14	236379	5.44
PEM	2	A	964329	964326	140.9	50010	5.19
PEM	2	B	3500538	3500530	140.41	178707	5.11
PEM	2	C	3616630	3616629	139.11	194141	5.37
PEM	2	D	2809090	2809089	140.73	131069	4.67
PEM	6	A	3097910	3097910	143.69	175117	5.65
PEM	6	B	3088734	3088732	142.51	167084	5.41
PEM	6	C	2599404	2599404	142.69	142123	5.47
PEM	6	D	2768442	2768440	143.16	139282	5.03
PEM	8	A	5159491	5159489	143.5	333238	6.46

PEM	8	B	4612154	4612148	141.79	303190	6.57
PEM	8	C	4550749	4550744	140.67	292525	6.43
PEM	8	D	5463253	5463251	142.01	348855	6.39
PSS	10	A	2925330	2925330	143.02	163075	5.57
PSS	10	B	4058362	4058362	142.13	234125	5.77
PSS	10	C	3846177	3846176	141.89	205618	5.35
PSS	10	D	3336661	3336660	143.24	193127	5.79
PSS	3	A	3433028	3433028	140.74	221344	6.45
PSS	3	B	3964611	3964609	137.68	231243	5.83
PSS	3	C	3569546	3569546	142.36	222676	6.24
PSS	3	D	2451038	2451038	143.44	156531	6.39
PSS	5	A	3548479	3548468	143.12	189748	5.35
PSS	5	B	3651930	3651927	142.89	201313	5.51
PSS	5	C	3101292	3101291	143.57	184545	5.95
PSS	5	D	2476940	2476938	143.75	142442	5.75
PFO	1	A	4579171	4579170	142.66	259173	5.66

PFO	1	B	6170088	6170081	141.84	320440	5.19
PFO	1	C	4673315	4673309	143.1	278145	5.95
PFO	1	D	3407073	3407068	143.83	184189	5.41
PFO	4	A	3537276	3537276	142.2	226648	6.41
PFO	4	B	5977871	5977869	139.51	358267	5.99
PFO	4	C	3111631	3111631	139.07	188950	6.07
PFO	4	D	3265825	3265815	140.66	201016	6.16
PFO	9	A	4593370	4593358	140.35	264361	5.76
PFO	9	B	4793269	4793267	141.58	308073	6.43
PFO	9	C	4412666	4412662	143.19	244637	5.54
PFO	9	D	4147902	4147898	142.07	208950	5.04
U	11	A	3966030	3966028	140.92	220435	5.56
U	11	B	3931483	3931481	140.05	224165	5.7
U	11	C	3524693	3524693	142.02	194443	5.52
U	11	D	2927945	2927945	141.09	138998	4.75
U	12	B	3157886	3157886	141.51	203353	6.44

U	12	C	3098881	3098880	141.02	189128	6.1
U	12	X	3367163	3367163	142.59	191614	5.69
U	12	Y	2781799	2781792	141.93	153918	5.53
U	7	A	3384627	3384626	139.84	197900	5.85
U	7	B	9485775	9485771	140.83	631634	6.66
U	7	C	3712324	3712318	143.23	215067	5.79
U	7	D	3084242	3084242	142.21	160226	5.19

BIBLIOGRAPHY

[1] S. Andrews. FastQC: a quality control tool for high throughput sequence data available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. 2010.

[2] Sarah M. Bisbing, David J. Cooper, David V. D'Amore, and Kristin N. Marshall. Determinants of conifer distributions across peatland to forest gradients in the coastal temperate rainforest of southeast Alaska: Conifer Distributions Across Hydrologic Gradients in the Temperate Rainforest. *Ecohydrology*, 9(2):354–367, March 2016.

[3] Sarah M. Bisbing and David V. DAmore. Nitrogen dynamics vary across hydrologic gradients and by forest community composition in the perhumid coastal temperate rainforest of southeast Alaska. *Canadian Journal of Forest Research*, 48(2):180–191, February 2018.

[4] Benjamin Buchfink, Chao Xie, and Daniel H. Huson. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1):59–60, January 2015.

[5] F. S. Chapin, A. D. Mcguire, J. Randerson, R. Pielke, D. Baldocchi, S. E. Hobbie, N. Roulet, W. Eugster, E. Kasischke, E. B. Rastetter, S. A. Zimov, and S. W. Running. Arctic and boreal ecosystems of western North America as components of the climate system. *Global Change Biology*, 6(S1):211–223, December 2000.

[6] M. R. Chivers, M. R. Turetsky, J. M. Waddington, J. W. Harden, and A. D. McGuire. Effects of Experimental Water Table and Temperature Manipulations on Ecosystem CO2 Fluxes in an Alaskan Rich Fen. *Ecosystems*, 12(8):1329–1342, December 2009.

[7] J. M. Clark, D. Ashley, M. Wagner, P. J. Chapman, S. N. Lane, C. D. Evans, and A. L. Heathwaite. Increased temperature sensitivity of net DOC production from ombrotrophic peat due to water table draw-down. *Global Change Biology*, 15(4):794–807, April 2009.

[8] Lewis M. Cowardin, Virginia Carter, Francis C. Golet, and Edward T. Laroe. Classification of Wetlands and Deepwater Habitats of the United States. In Jay H. Lehr and Jack Keeley, editors, *Water Encyclopedia*. John Wiley & Sons, Inc., Hoboken, NJ, USA, July 2005.

[9] Eric A. Davidson and Ivan A. Janssens. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature*, 440(7081):165–173, March 2006.

[10] Helen Decleyre, Kim Heylen, Tytgat Bjorn, and Anne Willems. Highly diverse nirK genes comprise two major clades that harbour ammonium-producing denitrifiers. *BMC Genomics*, 17, February 2016.

[11] Laure Dutaur and Louis V. Verchot. A global inventory of the soil CH4 sink. *Global Biogeochemical Cycles*, 21(4), December 2007.

[12] Lukasz Dziewit, Adam Pyzik, Krzysztof Romaniuk, Adam Sobczak, Pawel Szczesny, Leszek Lipinski, Dariusz Bartosik, and Lukasz Drewniak. Novel molecular markers for the detection of methanogens and phylogenetic analyses of methanogenic communities. *Frontiers in Microbiology*, 6, July 2015.

[13] David V. DAmore, Rick T. Edwards, and Frances E. Biles. Biophysical controls on dissolved organic carbon concentrations of Alaskan coastal temperate rainforest streams. *Aquatic Sciences*, 78(2):381–393, April 2016.

[14] David V. DAmore, Kiva L. Oken, Paul A. Herendeen, E. Ashley Steel, and Paul E. Hennon. Carbon accretion in unthinned and thinned young-growth forest stands of the Alaskan perhumid coastal temperate rainforest. *Carbon Balance and Management*, 10(1):25, October 2015.

[15] David V. DAmore, Chien-Lu Ping, and Paul A. Herendeen. Hydromorphic Soil Development in the Coastal Temperate Rainforest of Alaska. *Soil Science Society of America Journal*, 79(2):698, 2015.

[16] Jason B. Fellman, David V. DAmore, Eran Hood, and Pat Cunningham. Vulnerability of wetland soil carbon stocks to climate warming in the perhumid coastal temperate rainforest. *Biogeochemistry*, 133(2):165–179, April 2017.

[17] Mark E. Fenn, Mark A. Poth, John D. Aber, Jill S. Baron, Bernard T. Bormann, Dale W. Johnson, A. Dennis Lemly, Steven G. McNulty, Douglas F. Ryan, and Robert Stottlemyer. Nitrogen Excess in North American Ecosystems: Predisposing Factors, Ecosystem Responses, and Management Strategies. *Ecological Applications. 8(3). 1998. pp. 706-733*, 1998.

[18] Mary Firestone and Eric A. Davidson. Microbiological Basis of NO and

N2o Production and Consumption in Soil. In *Exchange of trace gases between terrestrial ecosystems and the atmoshpere.*, pages 7 – 21. Wiley, 1989.

[19] Scott M. Gende, Amy E. Miller, and Eran Hood. The effects of salmon carcasses on soil nitrogen pools in a riparian forest of southeastern Alaska. *Canadian Journal of Forest Research*, 37(7):1194–1202, 2007.

[20] Christos Gougoulias, Joanna M Clark, and Liz J Shaw. The role of soil microbes in the global carbon cycle: tracking the below-ground microbial processing of plant-derived carbon for manipulating carbon dynamics in agricultural systems. *Journal of the Science of Food and Agriculture*, 94(12):2362–2371, September 2014.

[21] Peter Groffman, Elisabeth Holland, David D. Myrold, G Philip Robertson, Xiaoming Zou, David Coleman, C.S. Bledsoe, and Phillip Sollins. Denitrification. In *Standard Soil Methods for Long Term Ecological Research*, pages 272–288. January 1999.

[22] Peter M. Groffman, Mark A. Altabet, J. K. Bhlke, Klaus Butterbach-Bahl, Mark B. David, Mary K. Firestone, Anne E. Giblin, Todd M. Kana, Lars Peter Nielsen, and Mary A. Voytek. Methods for Measuring Denitrification: Diverse Approaches to a Difficult Problem. *Ecological Applications*, 16(6):2091–2122, December 2006.

[23] Huazhi Han, James Hemp, Laura A. Pace, Hanlin Ouyang, Krithika Ganesan, Jung Hyeob Roh, Fevzi Daldal, Steven R. Blanke, and Robert B. Gennis. Adaptation of aerobic respiration to low O2 environments. *Proceedings of the National Academy of Sciences*, 108(34):14109–14114, August 2011.

[24] P.J. Hanson, N.T. Edwards, C.T. Garten, and J.A. Andrews. Separating root and soil microbial contributions to soil respiration: A review of methods and observations. *Biogeochemistry*, 48(1):115–146, January 2000.

[25] Robbie A. Hember. Spatially and temporally continuous estimates of annual total nitrogen deposition over North America, 18602013. *Data in Brief*, 17:134–140, January 2018.

[26] David J. Janetski, Dominic T. Chaloner, Scott D. Tiegs, and Gary A. Lamberti. Pacific salmon effects on stream ecosystems: a quantitative synthesis. *Oecologia*, 159(3):583–595, March 2009.

[27] A. R. Keyser, J. S. Kimball, R. R. Nemani, and S. W. Running. Simulating the

effects of climate change on the carbon balance of North American high-latitude forests. *Global Change Biology*, 6(S1):185–195, December 2000.

[28] Daehwan Kim, Li Song, Florian P. Breitwieser, and Steven L. Salzberg. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Research*, October 2016.

[29] Jiwoong Kim, Min Soo Kim, Andrew Y. Koh, Yang Xie, and Xiaowei Zhan. FMAP: Functional Mapping and Analysis Pipeline for metagenomics and metatranscriptomics studies. *BMC Bioinformatics*, 17, October 2016.

[30] Rex B. Kline. Assumptions in structural equation modeling. In *Handbook of structural equation modeling*, pages 111–125. Guilford Press, New York, NY, US, 2012.

[31] Beate Kraft, Marc Strous, and Halina E. Tegetmeyer. Microbial nitrate respiration Genes, enzymes and environmental distribution. *Journal of Biotechnology*, 155(1):104–117, August 2011.

[32] Daniel R. Lammel, Brigitte J. Feigl, Carlos C. Cerri, and Klaus Nsslein. Specific microbial gene abundances and soil parameters contribute to C, N, and greenhouse gas process rates after land use change in Southern Amazonian Soils. *Frontiers in Microbiology*, 6, October 2015.

[33] Daniel A. Lashof and Dilip R. Ahuja. Relative contributions of greenhouse gas emissions to global warming. *Nature*, 344(6266):529–531, April 1990.

[34] David J. Levy-Booth, Cindy E. Prescott, and Susan J. Grayston. Microbial functional genes involved in nitrogen fixation, nitrification and denitrification in forest ecosystems. *Soil Biology and Biochemistry*, 75:11–25, August 2014.

[35] Harris A. Lewin, Gene E. Robinson, W. John Kress, William J. Baker, Jonathan Coddington, Keith A. Crandall, Richard Durbin, Scott V. Edwards, Flix Forest, M. Thomas P. Gilbert, Melissa M. Goldstein, Igor V. Grigoriev, Kevin J. Hackett, David Haussler, Erich D. Jarvis, Warren E. Johnson, Aristides Patrinos, Stephen Richards, Juan Carlos Castilla-Rubio, Marie-Anne van Sluys, Pamela S. Soltis, Xun Xu, Huanming Yang, and Guojie Zhang. Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences*, 115(17):4325–4333, April 2018.

[36] Wei Li, Hui Li, Yong-di Liu, Ping Zheng, and James P. Shapleigh. Salinity-Aided Selection of Progressive Onset Denitrifiers as a Means of Providing Nitrite for Anammox. *Environmental Science & Technology*, 52(18):10665–10672, September 2018.

[37] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10–12, May 2011.

[38] Nicholas J. Morley, David J. Richardson, and Elizabeth M. Baggs. Substrate Induced Denitrification over or under Estimates Shifts in Soil N2/N2o Ratios. *PLoS ONE*, 9(9), September 2014.

[39] Nigel J. Mouncey and Samuel Kaplan. Oxygen Regulation of the ccoN Gene Encoding a Component of the cbb 3 Oxidase inRhodobacter sphaeroides 2.4.1t: Involvement of the FnrL Protein. *Journal of Bacteriology*, 180(8):2228–2231, April 1998.

[40] Cornelius Oertel, Jrg Matschullat, Kamal Zurba, Frank Zimmermann, and Stefan Erasmi. Greenhouse gas emissions from soilsA review. *Chemie der Erde - Geochemistry*, 76(3):327–352, October 2016.

[41] Brian D. Ondov, Todd J. Treangen, Pll Melsted, Adam B. Mallonee, Nicholas H. Bergman, Sergey Koren, and Adam M. Phillippy. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology*, 17:132, June 2016.

[42] T.B. Parkin and R.T. Venterea. Chapter 3. Chamber-Based Trace Gas Flux Measurements. In *Sampling Protocols.*, pages 3.1–3.39. 2010.

[43] Manuela M Pereira, Margarida Santana, and Miguel Teixeira. A novel scenario for the evolution of haemcopper oxygen reductases. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1505(2):185–208, June 2001.

[44] Manuela M. Pereira and Miguel Teixeira. Proton pathways, ligand binding and dynamics of the catalytic site in haem-copper oxygen reductases: a comparison between the three families. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1655:340–346, April 2004.

[45] Dorthe Groth Petersen, Steven J. Blazewicz, Mary Firestone, Donald J. Herman, Merritt Turetsky, and Mark Waldrop. Abundance of microbial genes associated with nitrogen cycling as indices of biogeochemical process rates across a vegetation gradient in Alaska. *Environmental Microbiology*, 14(4):993–1008, April 2012.

[46] Markus Reichstein, Ana Rey, Annette Freibauer, John Tenhunen, Riccardo Valentini, Joao Banza, Pere Casals, Yufu Cheng, Jose M. Grnzweig, James Irvine, Richard Joffre, Beverly E. Law, Denis Loustau, Franco Miglietta, Walter Oechel, Jean-Marc Ourcival, Joao S. Pereira, Alessandro Peressotti, Francesca Ponti, Ye Qi, Serge Rambal, Mark Rayment, Joan Romanya, Federica Rossi, Vanessa Tedeschi, Giampiero Tirone, Ming Xu, and Dan Yakir. Modeling temporal and large-scale spatial variability of soil respiration from soil water availability, temperature and vegetation productivity indices. *Global Biogeochemical Cycles*, 17(4), December 2003.

[47] G. P. Robertson, editor. *Standard soil methods for long-term ecological research*. Number 2 in Long-term ecological research network series. Oxford University Press, New York, 1999.

[48] Edward A. G. Schuur, Jason G. Vogel, Kathryn G. Crummer, Hanna Lee, James O. Sickman, and T. E. Osterkamp. The effect of permafrost thaw on old carbon release and net carbon exchange from tundra. *Nature*, 459(7246):556–559, May 2009.

[49] James P. Shapleigh. Denitrifying Prokaryotes. In Eugene Rosenberg, Edward F. DeLong, Stephen Lory, Erko Stackebrandt, and Fabiano Thompson, editors, *The Prokaryotes: Prokaryotic Physiology and Biochemistry*, pages 405–425. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[50] Filipa L. Sousa, Renato J. Alves, Jos B. Pereira-Leal, Miguel Teixeira, and Manuela M. Pereira. A Bioinformatics Classifier and Database for Heme-Copper Oxygen Reductases. *PLOS ONE*, 6(4):e19117, April 2011.

[51] Filipa L. Sousa, Renato J. Alves, Miguel A. Ribeiro, Jos B. Pereira-Leal, Miguel Teixeira, and Manuela M. Pereira. The superfamily of hemecopper oxygen reductases: Types and evolutionary considerations. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1817(4):629–637, April 2012.

[52] Courtney Sparacino-Watkins, John F. Stolz, and Partha Basu. Nitrate and periplasmic nitrate reductases. *Chemical Society reviews*, 43(2):676–706, January 2014.

[53] Daan R. Speth and Victoria J. Orphan. Metabolic marker gene mining provides insight in global mcrA diversity and, coupled with targeted genome reconstruction, sheds further light on metabolic potential of the Methanomassiliicoccales. *PeerJ*, 6:e5614, September 2018.

[54] J. M. Stafford, G. Wendler, and J. Curtis. Temperature and precipitation

of Alaska: 50 year trend analysis. *Theoretical and Applied Climatology*, 67(1-2):33–44, October 2000.

[55] Baris E. Suzek, Yuqi Wang, Hongzhan Huang, Peter B. McGarvey, and Cathy H. Wu. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, 31(6):926–932, March 2015.

[56] Leho Tedersoo, Mohammad Bahram, Sergei Plme, Sten Anslan, Taavi Riit, Urmas Kljalg, R. Henrik Nilsson, Falk Hildebrand, and Kessy Abarenkov. Response to Comment on Global diversity and geography of soil fungi. *Science*, 349(6251):936–936, August 2015.

[57] Rudolf K. Thauer, Anne-kristin Kaster, Henning Seedorf, Wolfgang Buckel, and Reiner Hedderich. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nature Reviews. Microbiology; London*, 6(8):579–91, August 2008.

[58] Jodie B. Ullman. Structural Equation Modeling: Reviewing the Basics and Moving Forward. *Journal of Personality Assessment*, 87(1):35–50, July 2006.

[59] Tadd A. Wheeler and Kathleen L. Kavanagh. Soil biogeochemical responses to the deposition of anadromous fish carcasses in inland riparian forests of the Pacific Northwest, USA. *Canadian Journal of Forest Research*, 47(11):1506–1516, November 2017.

[60] Tana E. Wood, Matteo Detto, and Whendee L. Silver. Sensitivity of Soil Respiration to Variability in Soil Moisture and Temperature in a Humid Tropical Forest. *PLOS ONE*, 8(12):e80965, December 2013.